

22  
Suggs, Faye (ASRC)

30f3

**From:** ANH LY [anh.ly@uspto.gov]  
**Sent:** Tuesday, August 07, 2007 10:29 AM  
**To:** STIC-EIC2100  
**Subject:** Database Search Request, Serial Number: 10/086,026

233580

**Requester:**  
ANH LY (P/2162)  
**Art Unit:**  
GROUP ART UNIT 2162  
**Employee Number:**  
77831  
**Office Location:**  
RND 03A39  
**Phone Number:**  
(571)272-4039  
**Mailbox Number:**

**Case serial number:**  
10/086,026  
**Class / Subclass(es):**  
707/1  
**Earliest Priority Filing Date:**  
02/26/2002  
**Format preferred for results:**  
Paper

**Search Topic Information:**

computing a segment-level expected usage value for each of the one or more word combinations in accordance with  $S(w_i) \times S(w_j) \times \dots \times S(w_m) / N(m-1)$  where "m" represents the number of words in the word combination, "N" represents the number of segments in the data corpus, and  $S(w)$  represents the number of unique segments in the data corpus that word  $w_i$  of the word combination is in and selecting for display another topic, said another topic associated with the most identified stored data items not associated with a previously identified display topic, wherein this step is repeated until all identified stored items in the result set have been accounted for.

**Special Instructions and Other Comments:**

# STIC Search Results Feedback Form

**EIC 2100**

Questions about the scope or the results of the search? Contact *the EIC searcher or contact:*

Alyson Dill, EIC 2100 Team Leader  
272-3527, RND 4B28

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 2133

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(Journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

**Comments:**

Drop off or send completed forms to STIC/EIC2100 RND, 4B28

Set	Items	Description
S1	3234483	WORD??? OR PHRAS? OR CLAUSE? OR CHARACTER? OR TERM? ? OR T-EXT? ? OR PART? ?(1W)SPEECH OR ALPHABET? OR SPELLING? OR LETTER?
S2	158312	S1(3N)(SET? ? OR GROUP? OR CLUSTER? OR BUNCH? OR COLLECTION? OR COMBINATION? OR AGGREGAT? OR ORDER? OR PATTERN? OR SEQUENCE? OR STRING?)
S3	9108181	(FRACTION? OR PART??? OR PORTION? OR SUBSET? OR FRAGMENT? - OR PIECE? OR SEGMENT? OR DIVID? OR DISAGGREGAT? OR SUBPART? OR PARTIT? OR PARTIAL? OR SUB)
S4	345315	S3(3N)(LEVEL? OR STEP? OR STAGE?) OR HIERARCH? OR PROGRESS-IV? OR RECURS? OR MULTILEVEL? OR MULTISTEP? OR MULTISTAG?
S5	96025	(WEIGH? OR IMPORT? OR RELEVAN? OR WORTH? OR SIGNIFICA? OR -VALUE?) (2N)(USAGE OR UTILI? OR EMPLOY? OR DEPLOY? OR USE? ? OR IMPLEMENT?)
S6	452290	S1:S4(5N)(SENSE? OR USEFUL? OR PRIORIT? OR GRAVITY OR POINT? ? OR INFLUEN? OR STRESS? OR RANK??? ? OR SALIEN? OR NOTEWORTH? OR BEARING? OR CONCERN? OR GIST? ? OR INTEREST?)
S7	25418	S5:S6(7N)(MONITOR? OR EXAMIN? OR DETECT? OR UNCOVER? OR REVEAL? OR ASSESS? OR EVALUAT? OR INSPECT?)
S8	40514	S5:S6(7N)(DETERMIN? OR COMPAR? OR DISCERN? OR ASCERTAIN? OR ANALY? OR IDENT? OR CHECK? OR VERIF? OR JUDG???)
S9	7639	(DETERMIN? OR PICK??? OR CHOOS? OR SELECT? OR CHOSEN OR SPECIF? OR DESIGNAT? OR INDICAT?) (5N)(TOPIC? OR SUBJECT?()) HEAD-ING OR INDEX?(2N)TERM? OR KEYWORD? OR KEYPHRASE? OR KEYCLAUSE? OR KEY() (WORD? OR PHRASE? OR CLAUSE?)
S10	12	S2 AND S4 AND S5 AND S7:S8 AND S9
S11	78	S2 AND S4 AND S9
S12	66	S11 NOT S10
S13	15	S12 AND RETRIEV?(3N)(SYSTEM? ? OR PROCESS?)

File 350:Derwent WPIX 1963-2007/UD=200754

(c) 2007 The Thomson Corporation

File 347:JAPIO Dec 1976-2007/Mar(Updated 070809)

(c) 2007 JPO & JAPIO

10/69,K/2 (Item 2 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2007 The Thomson Corporation. All rts. reserv.

0014423099 - Drawing available  
WPI ACC NO: 2004-613245/200459  
XRPX Acc No: N2004-484638

**Search engine computer system, has natural language processor in search component that translates search terms received by user interface into prioritized clustered tokens**

Patent Assignee: MICROSOFT CORP (MICT)  
Inventor: JAYANTI H; KONASEWICH P; STUMPF M D

**Patent Family** (1 patents, 1 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
US 6775666	B1	20040810	US 2001867228	A	20010529	200459 B

Priority Applications (no., kind, date): US 2001867228 A 20010529

#### Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
US 6775666	B1	EN	30	20	

#### Alerting Abstract US B1

NOVELTY - A user interface receives user-defined search terms in searchable content database including an index from an information source, and displays the results. A search component searches the terms and retrieves the information containing the search term. A natural language processor in the search component translates the search **terms** into prioritized **clustered** tokens.

DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

- 1.method for searching and retrieving information; and
- 2.computer readable medium storing information searching program.

USE - For index searching queries in computer system e.g. Internet.

ADVANTAGE - Users can be directed to general or specific content within an article outline and related articles. Multiple query styles can be searched to find **relevant** matches. **User** queries are **analyzed to determine most /less important elements** , and can be formed in an ad-hoc, free-form manner. Treatment of **hierarchical** index data can be combined **with** a natural language processor to provide more accurate and detailed access to indexed content. Retrieves search results within less processing time.

DESCRIPTION OF DRAWINGS - The figure shows the architecture of the computer system.

**Title Terms/Index Terms/Additional Words:** SEARCH; ENGINE; COMPUTER; SYSTEM; NATURAL; LANGUAGE; PROCESSOR; COMPONENT; TRANSLATION; TERM; RECEIVE; USER ; INTERFACE; CLUSTER; TOKEN

#### Class Codes

International Classification (Main): G06F-017/30  
US Classification, Issued: 707005000, 707004000

File Segment: EPI;

DWPI Class: T01

Manual Codes (EPI/S-X): T01-J16C3; T01-J16C6; T01-N03A2; T01-S03

**Alerting Abstract** ...containing the search term. A natural language processor in the search component translates the search **terms** into prioritized **clustered** tokens....within an article outline and related articles. Multiple query styles can be searched to find **relevant** matches. User queries are **analyzed to determine most /less important elements** , and **can** be formed in an ad-hoc, free-form manner. Treatment of **hierarchical** index data can be combined **with** a natural language processor to provide more accurate and detailed access to indexed content. Retrieves...

#### **Original Publication Data by Authority**

##### **Original Abstracts:**

A method and system for searching index databases allows a user to search for **specific** information using high- **level key words** , questions, or **sentences** . The system includes three main segments: a searchable content database, a run time search component, and...

...exact match search, a natural language processor (NLP), and a full text search. Indexes, prioritized **search** tokens, and **word clusters** are combined **to create** a better search experience. A user's query is processed into prioritized clustered tokens using the NLP, token priority rules, and **word clusters**.

##### **Claims:**

...the search terms, the search component comprising a natural language processor for translating the search **terms** into prioritized **clustered** tokens.

# \* YOUR ASSIGNEE \*

10/69,K/4 (Item 4 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2007 The Thomson Corporation. All rts. reserv..

0013616347 - Drawing available  
WPI ACC NO: 2003-711698/200367  
Related WPI Acc No: 2006-087806  
XRPX Acc No: N2003-569242

**Data corpus topics identifying method for information retrieval systems, involves designating word combination as topic if determined segment-level actual usage value of combination is greater than expected usage value**

Patent Assignee: PLIANT TECHNOLOGIES INC (PLIA-N)  
Inventor: AKILESWAR S; CHILDERS R; KOTLAR D; ODOM P S  
Patent Family (1 patents, 1 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
US 20030167252	A1	20030904	US 200286026	A	20020226	200367 B

Priority Applications (no., kind, date): US 200286026 A 20020226

## Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
US 20030167252	A1	EN	15	8	

### Alerting Abstract US A1

NOVELTY - The method involves determining a segment - level actual usage value for word combinations. A segment - level expected usage value is computed for each word combination. A word combination is designated as a topic if the segment - level actual usage value of the combination is greater than the segment - level expected usage value. One word in each word combination is selected from a set of word lists.

DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

- 1.a program storage device to identify topics in a data corpus
- 2.a method to display a list of topics associated with data items stored in a database.

USE - Used for identifying topics in a data corpus of information retrieval systems.

ADVANTAGE - The method provides a fast, low cost and automated classification of large amounts of data that is consistent with the semantic content of the data. The method provides a collection of topics that are used to guide information retrieval and the display of topic classifications during user query operations.

DESCRIPTION OF DRAWINGS - The drawing shows a flow chart representing the method to identify topics in a corpus of data.

**Title Terms/Index Terms/Additional Words:** DATA; CORPUS; TOPIC; IDENTIFY; METHOD; INFORMATION; RETRIEVAL; SYSTEM; DESIGNATED; WORD; COMBINATION; DETERMINE; SEGMENT; LEVEL; ACTUAL; VALUE; GREATER

## Class Codes

International Classification (Main): G06F-017/30  
US Classification, Issued: 707001000

File Segment: EPI;  
DWPI Class: T01  
Manual Codes (EPI/S-X): T01-J05B; T01-S03

Data corpus topics identifying method for information retrieval systems, involves designating word combination as topic if determined segment-level actual usage value of combination is greater than expected usage value

...NOVELTY - The method involves determining a segment - level actual usage value for word combinations . A segment - level expected usage value is computed for each word combination . A word combination is designated as a topic if the segment - level actual usage value of the combination is greater than the segment - level expected usage value . One word in each word combination is selected from a set of word lists.

Original Publication Data by Authority

#### Original Abstracts:

A technique to determine topics associated with, or classifications for, a data corpus uses an initial domain-specific word list to identify word combinations (one or more words) that appear in the data corpus significantly more often than expected. Word combinations so identified are selected as topics and associated with a user-specified level of granularity. For example, topics may be associated with each table entry, each image, each sentence, each paragraph, or an entire file. Topics may...

#### Claims:

...claimed is:<b>1</b>. A method to identify topics in a data corpus having a plurality of segments , comprising: determining a segment-level actual usage value for one or more word combinations; computing a segment- level expected usage value for each of the one or more word combinations ; and designating a word combination as a topic if the segment - level actual usage value of the word combination is substantially greater than the segment - level expected usage value of the word combination.>

10/69,K/5 (Item 5 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2007 The Thomson Corporation. All rts. reserv.

0013463742 - Drawing available  
WPI ACC NO: 2003-555262/200352  
XRPX Acc No: N2003-440982

**Electronic document file locating, ranking and marking method in internet, involves ordering group of pie charts that represent collection of retrieved documents, hierarchically based on relevance of common key word**  
Patent Assignee: ARAHA INC (ARAH-N)  
Inventor: HUSSAM A A

**Patent Family** (1 patents, 1 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
US 20030050927	A1	20030313	US 2001318168	P	20010907	200352 B
			US 2002127638	A	20020422	

Priority Applications (no., kind, date): US 2001318168 P 20010907; US 2002127638 A 20020422

#### Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
US 20030050927	A1	EN	54	48	Related to Provisional US 2001318168

#### Alerting Abstract US A1

NOVELTY - A subset of electronic document files having a **selected** common **key word** is retrieved from a universe of electronic document files. The key word is marked with a color highlighter. A group of perceptible pie charts corresponding to documents is provided. The pie charts that represent the collection of retrieved documents are ordered, **hierarchically** based on relevance of key word.

DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

- 1.method of providing abstract visual representations of a desired subset of data derived from a set of data;
- 2.system of organizing an arbitrary collection of electronic documents;
- 3.method of extracting and arranging a subset of electronic documents from a larger group of electronic documents; method of information gathering and encoding;
- 4.method of organizing and sharing electronic document files among multiple users; and
- 5.method of information acquisition.

USE - For locating, ranking and marking electronic document including hypertext markup language (HTML) documents in internet or intranet.

ADVANTAGE - The semantic highlighting enhances the rate at which people locate and understand web-based documents is enhanced and allows for metadata that is not static to be created by the author or other users of the document. By using visual metadata in the form of pie charts, the user is allowed to perform rapid **assessment** of the **relevance** of documents **located** by the search engine.

DESCRIPTION OF DRAWINGS - The figure shows a flow chart illustrating a task analysis for locating and using a document.

**Title Terms/Index Terms/Additional Words:** ELECTRONIC; DOCUMENT; FILE; LOCATE; RANK; MARK; METHOD; ORDER; GROUP; PIE; CHART; REPRESENT; COLLECT;



RETRIEVAL; **HIERARCHY** ; BASED; RELEVANT; COMMON; KEY; WORD

**Class Codes**

International Classification (Main): G06F-017/30

US Classification, Issued: 707005000

File Segment: EPI;

DWPI Class: T01

Manual Codes (EPI/S-X): T01-J05B1; T01-N03A2; T01-N03B2

...method in internet, involves ordering group of pie charts that represent collection of retrieved documents, hierarchically based on relevance of common key word

**Alerting Abstract** ...NOVELTY - A subset of electronic document files having a **selected common key word** is retrieved from a universe of electronic document files. The key word is marked with...

...documents is provided. The pie charts that represent the collection of retrieved documents are ordered, **hierarchically** based on relevance of key word....visual metadata in the form of pie charts, the user is allowed to perform rapid **assessment** of the **relevance** of documents **located** by the search engine.

**Title Terms...**/Index Terms/Additional Words: **HIERARCHY** ;

**Original Publication Data by Authority**

**Original Abstracts:**

...within a universe of preexisting documents to extract a subset of relevant documents is disclosed. **The user selects** search terms or **key words** , and an **application program** performs a search of the universe of documents, compiles a subset or collection of documents based upon the search **terms** or **keywords** selected, and presents **the resulting collection** of documents to the user. An abstract marker such as a color highlighter, e.g...

**Claims:**

...each electronic document file within said subset of electronic document file with a first abstract **indicia** :providing a **group** of second abstract **indicias** each corresponding to a document in said subset of electronic documents, said second abstract **indicias** being perceptible; and ordering said group of abstract **indicias** **hierarchically** based upon the relevance of said characteristic.

10/69,K/6 (Item 6 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2007 The Thomson Corporation. All rts. reserv.

0012899069 - Drawing available  
WPI ACC NO: 2002-758658/200282  
XRPX Acc No: N2002-597251

**Text input method in virtual environment, involves monitoring positions of fingers by sensor glove that is calibrated with dynamic threshold values indicating occurrence of finger press**

Patent Assignee: UNIV NEW YORK STATE RES FOUND (UYN Y)

Inventor: EVANS F; SKIENA S; VARSHNEY A

**Patent Family** (1 patents, 1 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
US 6407679	B1	20020618	US 199894910	P	19980731	200282 B
			US 1999364433	A	19990730	

Priority Applications (no., kind, date): US 199894910 P 19980731; US 1999364433 A 19990730

#### Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
US 6407679	B1	EN	52	9	Related to Provisional US 199894910

#### Alerting Abstract US B1

NOVELTY - A sensor glove is calibrated by establishing threshold values, when a user enters a sample sequence. The positions of the fingers are monitored by the sensor glove. If a finger press has passed the threshold value, a feedback is provided to the user to indicate that a key is entered and the key is stored. Key words are separated by recognizing spaces in the stored sequence of keys and are matched with words in a dictionary, to select most probable word sequence.

USE - For entering text including Chinese symbols in a virtual environment created by computer system.

ADVANTAGE - Recognizes fine finger movements such as key entry on a virtual key board using low cost, low resolution sensor glove. Eliminates the inherent noise in the finger movement data by using a low-pass Gaussian filter.

DESCRIPTION OF DRAWINGS - The figure shows a flowchart explaining finger press recognition procedure.

**Title Terms/Index Terms/Additional Words:** TEXT; INPUT; METHOD; VIRTUAL; ENVIRONMENT; MONITOR; POSITION; FINGER; SENSE; GLOVE; CALIBRATE; DYNAMIC; THRESHOLD; VALUE; INDICATE; OCCUR; PRESS

#### Class Codes

International Classification (Main): H03M-011/00  
US Classification, Issued: 341020000, 400475000, 400479200, 345702000, 345811000, 345168000, 345773000

File Segment: EPI;  
DWPI Class: U21  
Manual Codes (EPI/S-X): U21-A05D

...NOVELTY - A sensor glove is calibrated by establishing threshold values, when a user enters a sample sequence. The positions of the fingers are monitored by the sensor glove...

...sequence of keys and are matched with words in a dictionary, to select most probable word sequence.

## Original Publication Data by Authority

### Claims:

...character for each finger press movement; calibrating the at least one glove by establishing threshold **values** through a **user** inputting a sample sequence, **said** threshold **values** **indicating** an occurrence of a finger press; **monitoring** the positions of the plurality of the fingers by the at least one sensor glove...

...matching the key words with one or more words in the dictionary, generating all possible **permutations** of **word sequences**, and **selecting** the most probable **word sequence** or partial sentence; when a partial sentence is selected, **generating feedback** to the user **concerning** the **selected partial sentence**; **returning** to said **monitoring** step; and **if** the user indicates the end of a sentence, **erasing** the stored **sequence** of keys, storing the last most **probable word sequence** as a sentence, and returning to said monitoring step.

10/69,K/7 (Item 7 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2007 The Thomson Corporation. All rts. reserv.

0012639968 - Drawing available  
WPI ACC NO: 2002-489085/200252  
XRPX Acc No: N2002-386613

Automatic summary generation method for text documents, involves  
calculating sum of scores of words and sentences extracted from document,  
using which top-ranked sentences and key word list are generated and output  
Patent Assignee: GUO Z L (GUOZ-I); INT BUSINESS MACHINES CORP (IBMC);  
YANG L P (YANG-I)

Inventor: GUO Z L; YANG L P

Patent Family (2 patents, 1 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
US 20020052901	A1	20020502	US 2001943341	A	20010831	200252 B
US 7017114	B2	20060321	US 2001943341	A	20010831	200621 E

Priority Applications (no., kind, date): US 2001943341 A 20010831; CN  
2000128668 A 20000907

#### Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
US 20020052901	A1	EN	7	2	

#### Alerting Abstract US A1

NOVELTY - A set of sentences and words are extracted from document by different processes. A score is set for each word and the sentence. The sum of the scores of words and that of sentences are calculated and if the calculated scores change apparently, the sum of the scores is computed again. The top-ranked sentences determined with respect to scores are output as summary and the top-ranked words are output as keyword list of the document.

DESCRIPTION - An INDEPENDENT CLAIM is included for a computer program product for automatically generating summaries for text documents.

USE - For generating summary for text documents automatically.

ADVANTAGE - A comprehensive summary including the important ideas of document is generated efficiently.

DESCRIPTION OF DRAWINGS - The figure shows a flowchart for automatic summary generation process.

Title Terms/Index Terms/Additional Words: AUTOMATIC; SUMMARY; GENERATE;  
METHOD; TEXT; DOCUMENT; CALCULATE; SUM; SCORE; WORD; SENTENCE; EXTRACT;  
TOP; RANK; KEY; LIST; OUTPUT

#### Class Codes

International Classification (Main): G06F-017/21

International Classification (+ Attributes)

IPC + Level Value Position Status Version

G06F-0015/00 A I F B 20060101

G06F-0017/00 A I L B 20060101

G06F-0017/21 A I L B 20060101

G06F-0017/30 A I L B 20060101

G06F-0007/00 A I L B 20060101

US Classification, Issued: 707531000, 707532000, 715531000, 715530000,  
707001000, 707003000, 707004000, 707006000

File Segment: EPI;

DWPI Class: T01

Manual Codes (EPI/S-X): T01-J11C; T01-S03

...NOVELTY - A **set** of sentences and **words** are extracted from document by different processes. A score is **set** for each **word** and the sentence. The sum of the scores of words and that of sentences are...

#### Original Publication Data by Authority

#### Original Abstracts:

A method and program product to generate summaries for text documents. A user can also **specify** a query, topic, and **terms** that he/she is **interested** in. This method **determines** the importance of **each** sentence by using the linguistic **salience** of the **word** to the user **profile**, the similarity among the word, the query and topic provided by a user and the...

...word, this method computes the score for each sentence in the set of sentences according to the score of **words** composing it and the position of the sentence in a section and a paragraph...

...A method and program product to generate summaries for text documents. A user can also **specify** a query, topic, and **terms** that he/she is **interested** in. This method **determines** the importance of each **sentence** by using the linguistic **salience** of the **word** to the user profile, the **similarity** among the **word**, the query and topic provided by a user and the sum of scores of the...

...After computing the score for each word, this method computes the score for each sentence in the **set** of sentences according to the score of **words** **composing** it and the position of the sentence in a section and a paragraph.

#### Claims:

...<b>1</b>. An automatic method for generating summaries for text documents, comprising steps of: generating a **set** of **sentences** for a **set** of documents by document discourse analysis and a **set** of **words** by morphologic process; initializing a score for each **word** in the **set** of **words** and for each sentence in the **set** of sentences; computing the score for each **word** in the **set** of **words** according to the score of sentences containing it and the correlation degree **between** the **word** and the user information; computing the score for each sentence in the **set** of **sentences** according to the score of words composing it and the position of the sentence in...

...a set of sentences for a set of documents by document discourse analysis and a **set** of **words** by morphologic process; initializing a word score for **each** **word** in the **set** of **words**, a sentence **score** for each sentence in the set of sentences and a score sum; computing an **aggregated** **word** score for said each **word** according to an **aggregate** of sentence **scores** of sentences **containing** said **each** **word** and to a degree of correlation between said **each** **word** and user related information; wherein said **aggregated** **word** **score** ( $SCORE[w]$ ) has a weighted ( $\lambda$ ) relationship with each of said aggregated sentence score ( $SCORE[s]$ ), linguistic salience of said each word to a user profile (salience( $w$ , user summarization profile)), similarities among said **each** **word**, a **query** and a provided topic (salience( $w$ , user's query or topic)), similarities among said **each** **word** and **terms** in titles of the documents (salience( $w$ , title words)), a ratio of an occurrence number for said each word in a document to a total occurrence number for said **each** **word** in the **set** of documents ( $FREQUENCY(w/d)/FREQUENCY(w/D)$ ), and a ratio of a

number of documents including said each word to a total number of documents in the **set** of documents ( $\text{NUMBER}(d, d\&\#xE800;w) / \text{NUMBER}(D)$ ), of the  
form  
 $\text{SCORE}[w] = \lambda_1 \cdot \text{saliency}(w, \text{usersummarizationprofile}) + \lambda_2 \cdot \text{saliency}(w, \text{user'squeryortopic}) + \lambda_3 \cdot \text{Sigma}(\text{SCORE} \dots$

$\dots w) / \text{NUMBER}(D)$ ; computing an aggregated sentence score for said each sentence according to an **aggregate** of **word** scores composing said each sentence and a respective **sentence position** in a section and a **paragraph**; **comparing** an **aggregate** sum with said score sum, said aggregate sum being a sum of **aggregated word** scores and **aggregated** sentence scores; and if said aggregate sum is different than said score sum, **returning** to the **step** of computing the **aggregated word** score; otherwise, outputting top-ranked sentences according to sentence score as a summary of the

10/69,K/8 (Item 8 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2007 The Thomson Corporation. All rts. reserv.

0010863004

WPI ACC NO: 2001-482041/200152

XRPX Acc No: N2001-356749

Document classification for information retrieval system, involves comparing created term and document vectors and storing document at location relative to category node with term vector with preset relevance ranking

Patent Assignee: SUN MICROSYSTEMS INC (SUNM)

Inventor: MOCKER J D; SNOW W A

Patent Family (1 patents, 1 countries)

Patent

Application

Number	Kind	Date	Number	Kind	Date	Update
US 6185550	B1	20010206	US 1997874783	A	19970613	200152 B

Priority Applications (no., kind, date): US 1997874783 A 19970613

#### Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
US 6185550	B1	EN	19	9	

#### Alerting Abstract US B1

NOVELTY - Term vectors containing weights assigned to each of one or more common terms in the corresponding terms file are created and are compared with created document vectors of a document to provide relevance ranking between the terms file and document. The document is stored at a location corresponding to category node having a term vector which has a relevance ranking that matches a selected criteria.

DESCRIPTION - A class **hierarchy** is created by providing several category nodes, each of which create term files. Class **hierarchy** having a root category node within a free data structure is initialized and displayed. User selected commands for manipulating the class **hierarchy** are entered. A category command is processed in response to the user selected command having predefined state which causes the class **hierarchy** to contain several category nodes. Category nodes include category name, node type, node ID, parent ID, link ID which are all stored in the database. When the node type is predefined type a new category node is allowed to be added to the selected category nodes, otherwise new category node is prevented from being added to the category nodes. The node ID defines the unique directory. The parent ID is indicating the node ID of a parent category node. The link ID is indicating the node ID of several category nodes when the node type is of a predetermined type. INDEPENDENT CLAIMS are also included for the following:

1.Document classifying;

2.Document classification program

USE - For classification of documents within defined categories using class **hierarchy** in information retrieval system.

ADVANTAGE - Since the automatic document classification within user defined categories is provided, the user can interactively search for documents according to search terms defined within user defined categories. Since documents are ranked according to **relevance** and a **user** specified number of documents **which** are most **relevant** are returned, multiple users can access the document via network.

DESCRIPTION OF DRAWINGS - The figure shows the flowchart of main procedure utilized in creation of the document directory **hierarchy**.

**Title Terms/Index Terms/Additional Words:** DOCUMENT; CLASSIFY; INFORMATION; RETRIEVAL; SYSTEM; COMPARE; TERM; VECTOR; STORAGE; LOCATE; RELATIVE; CATEGORY; NODE; PRESET; RELEVANT; RANK

**Class Codes**

International Classification (Main): G06F-017/30  
US Classification, Issued: 707001000, 707003000, 707005000, 707100000, 707102000, 707514000, 707907000  
File Segment: EPI;  
DWPI Class: T01  
Manual Codes (EPI/S-X): T01-E01C; T01-H07C5E; T01-J05B2B; T01-J05B3; T01-J05B4; T01-S03

**Original Titles:**

Method and apparatus for classifying documents within a class **hierarchy** creating term vector, term file and relevance ranking.

**Alerting Abstract DESCRIPTION** - A class **hierarchy** is created by providing several category nodes, each of which create term files. Class **hierarchy** having a root category node within a free data structure is initialized and displayed. User selected commands for manipulating the class **hierarchy** are entered. A category command is processed in response to the user selected command having predefined state which causes the class **hierarchy** to contain several category nodes. Category nodes include category name, node type, node ID, parent...

...USE - For classification of documents within defined categories using class **hierarchy** in information retrieval system.

...documents according to search terms defined within user defined categories. Since documents are ranked according to **relevance** and a user specified number of documents **which** are most **relevant** are returned, multiple users can access the document via network...

...The figure shows the flowchart of main procedure utilized in creation of the document directory **hierarchy** .

**Original Publication Data by Authority**

**Original Abstracts:**

A method for classifying a document based on content within a class **hierarchy** . The class **hierarchy** comprises a plurality of category nodes stored within a tree data structure. Each of the...

...includes a category name corresponding to a unique directory and a category definition comprising a **set** of defining **terms** . The class **hierarchy** is searched to determine appropriate categories for classification of the document. The document is then...

**Claims:**

A method for classifying a document based on content within a class **hierarchy** , the **method** comprising:initializing the class **hierarchy** , the class **hierarchy** having a root **category** node within a tree data structure, the root category node having a user-defined category name;displaying the class **hierarchy** ;accepting a user-selected command **for** manipulating the class **hierarchy** ;processing a category command in response to **the** user-selected command having a first predefined state, causing the class **hierarchy** to contain a plurality of category nodes, said processing **the** category command further comprising:storing a



category name in one of the plurality of category...

...plurality of category nodes when the nodetype is of a predefined type;creating a class **hierarchy** by providing a plurality of category nodes stored in a tree data structure within a memory, each of said plurality of **category** nodes having a category name corresponding to a unique directory and a **set** of defining **terms** ;creating a plurality of terms files, each of said plurality of terms files corresponding to one of said **plurality** of category **nodes** and including a corresponding **set** of **defining terms** and one or more document fragments stored under said one of said plurality of category nodes, said **set** of defining **terms** including a **term** corresponding to one of said plurality of category nodes and said one or more document fragments including a reference to **one** or more **documents** and **indexing** information **indicating** contiguous multi-term portions of said documents to be extracted during indexing, said **set** of defining **terms** and said document fragments together providing a definition of files to be contained in said unique directory referenced by said **one** of said **plurality** of category nodes;creating one or more term vectors for each of said terms files...

...vector containing a weight assigned to the terms of the document according to frequency of **occurrence** ;providing a relevance **ranking** between said **terms** files and said document by **comparing** said document vector with said one or more term vectors; andstoring said document within said document directory **hierarchy** at a location corresponding to a category node having a **term** vector which has a relevance **ranking** that matches a selected criteria.

10/69,K/9 (Item 9 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2007 The Thomson Corporation. All rts. reserv.

0010356345 - Drawing available  
WPI ACC NO: 2000-671975/200065  
XRPX Acc No: N2000-498142

**Computer implemented method for dynamic generation of hypertext links, involves selecting terms from user selected portion of source document and automatically creating hypertext link to target documents**

Patent Assignee: PERSPECTA INC (PERS-N)  
Inventor: HOROWITZ D M; RENNISON E F; RUFFLES J W; STRAUSFELD L S

**Patent Family** (1 patents, 1 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
US 6122647	A	20000919	US 199881695	A	19980519	200065 B

Priority Applications (no., kind, date): US 199881695 A 19980519

#### Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
US 6122647	A	EN	18	9	

#### Alerting Abstract US A

NOVELTY - A user selected portion (304) of the source document (300) is received and from which terms are selected. For a selected term, hypertext link to target document (310) is automatically created with relevance to the term. A menu of hypertext links to the target documents is created with selected term as link anchor.

DESCRIPTION - An INDEPENDENT CLAIM is also included for computer implement system for dynamically generating content dependent hypertext links.

USE - For dynamic generation of hypertext links from source document to target document in world wide web sites.

ADVANTAGE - By generating hypertext links dynamically, the document appropriate to user interests is customized effectively. The user has full control over the semantic content used in defining links. The static generation of tags permits links in the source document to be current at all times while giving the publisher editorial control over tags and keywords. The publisher is able to update the tags or links of the document effectively using automatic generation of links.

DESCRIPTION OF DRAWINGS - The figure shows illustration of navigation paradigm in hyper link generating method.

300 Source document  
304 User selected portion  
310 Target document

**Title Terms/Index Terms/Additional Words:** COMPUTER; IMPLEMENT; METHOD; DYNAMIC; GENERATE; LINK; SELECT; TERM; USER; PORTION; SOURCE; DOCUMENT; AUTOMATIC; TARGET

#### Class Codes

International Classification (Main): G06F-017/21  
US Classification, Issued: 707513000, 707003000, 707005000, 707514000, 707531000

File Segment: EPI;  
DWPI Class: T01  
Manual Codes (EPI/S-X): T01-H07C5E; T01-J11C1

**Original Publication Data by Authority**

**Original Abstracts:**

A system, method, and software product create contextual hypertext links **relevant** to a **user** selected portion of a source document. The contextual links enable the user to dynamically associate...

...and the target document when the source document was created. The method includes selecting terms **relevant** to the **user** selected portion by linguistic **analysis** which selects the most frequently occurring terms. From the selected terms target documents relevant to the selected terms are identified. The target documents are **selected** by identifying **topics** that are associated with, or described by, the selected terms. Contextual links are created between...

...the documents in the contextual links. The system includes a knowledge base of topics, including **hierarchical** relations between topics, and associations of topics and **terms**. A document **collection** includes documents and references to documents, and URL or other addressing information for the documents...

10/69,K/10 (Item 10 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2007 The Thomson Corporation. All rts. reserv.

0009501261 - Drawing available  
WPI ACC NO: 1999-443821/199937  
XRPX Acc No: N1999-331024

**Method for dynamic presentation of contents of several documents on display by receiving several documents and providing several capsule overviews corresponding to several documents**

Patent Assignee: APPLE COMPUTER INC (APPY); BELLAMY R K E (BELL-I);  
BOGURAEV B (BOGU-I); EMMA BELLAMY R K (BELL-I); WONG Y Y (WONG-I)  
Inventor: BELLAMY R K E; BOGURAEV B; EMMA BELLAMY R K; WONG Y Y

**Patent Family** (11 patents, 79 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
WO 1999026172	A1	19990527	WO 1998US24384	A	19981116	199937 B
AU 199914606	A	19990607	AU 199914606	A	19981116	199943 E
EP 951686	A1	19991027	EP 1998958594	A	19981116	199950 E
			WO 1998US24384	A	19981116	
US <sup>a</sup> 6353824	B1	20020305	US 1997972935	A	19971118	200224 E
US 20020133480	A1	20020919	US 1997972935	A	19971118	200264 E
			US 2001998406	A	20011129	
CA 2277209	C	20030114	CA 2277209	A	19981116	200309 E
			WO 1998US24384	A	19981116	
US 6553373	B2	20030422	US 1997972935	A	19971118	200330 E
			US 2001998406	A	20011129	
US 20030158843	A1	20030821	US 1997972935	A	19971118	200356 E
			US 2001998406	A	20011129	
			US 2003371430	A	20030220	
US 20040024747	A1	20040205	US 1997972935	A	19971118	200411 E
			US 2001953672	A	20010912	
US 6865572	B2	20050308	US 1997972935	A	19971118	200518 E
			US 2001998406	A	20011129	
			US 2003371430	A	20030220	
US 20050091591	A1	20050428	US 1997972935	A	19971118	200530 E
			US 2001998406	A	20011129	
			US 2003371430	A	20030220	
			US 2004972557	A	20041025	

Priority Applications (no., kind, date): US 2004972557 A 20041025; US 2003371430 A 20030220; US 2001998406 A 20011129; US 2001953672 A 20010912; US 1997972935 A 19971118

#### Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
WO 1999026172	A1	EN	53	8	
National Designated States,Original: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH GM HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG UZ VN YU ZW					
Regional Designated States,Original: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW					
AU 199914606	A	EN			Based on OPI patent WO 1999026172
EP 951686	A1	EN			PCT Application WO 1998US24384
					Based on OPI patent WO 1999026172
Regional Designated States,Original: DE FR GB					
US 20020133480	A1	EN			Continuation of application US 1997972935
					Continuation of patent US 6353824
CA 2277209	C	EN			PCT Application WO 1998US24384

US 6553373	B2	EN	Based on OPI patent WO 1999026172
1997972935			Continuation of application US
US 20030158843	A1	EN	Continuation of patent US 6353824
1997972935			Continuation of application US
2001998406			Continuation of application US
US 20040024747	A1	EN	Continuation of patent US 6353824
1997972935			Continuation of patent US 6553373
			Continuation of application US
US 6865572	B2	EN	Continuation of patent US 6353824
1997972935			Continuation of application US
2001998406			Continuation of application US
US 20050091591	A1	EN	Continuation of patent US 6353824
1997972935			Continuation of patent US 6553373
			Continuation of application US
2001998406			Continuation of application US
2003371430			Continuation of application US
			Continuation of patent US 6353824
			Continuation of patent US 6553373

#### **Alerting Abstract WO A1**

NOVELTY - Method presents documents thematic capsule overviews (206) to users derived for entire document showing core content of average length article in more accurate and representative manner than using conventional techniques. Overviews are delivered in a variety of presentation modes and allow users to quickly get the sense of what a document is about and decide if they want to be read in more detail.

USE - For reviewing documents and presenting them in a manner that allows the user to quickly ascertain their contents.

ADVANTAGE - User can decide whether he or she desires to be actively involved in the presentation.

DESCRIPTION OF DRAWINGS - The drawing shows a simple flow chart illustrating a method for the dynamic presentation of several documents. 206 displaying capsule overviews

**Title Terms/Index Terms/Additional Words:** METHOD; DYNAMIC; PRESENT; CONTENT ; DOCUMENT; DISPLAY; RECEIVE; CAPSULE; CORRESPOND

#### **Class Codes**

International Classification (Main): G06F-017/21, G06F-017/30, G06F-007/00  
 US Classification, Issued: 707001000, 707005000, 707003000, 715526000, 707005000, 707001000, 707003000, 707104100, 707501100, 707530000, 707531000, 704009000, 707005000, 707001000, 707003000, 707102000, 707104100, 707501100, 707530000, 707531000, 707513000, 704009000, 707005000, 707001000, 707003000, 707100000, 707104100, 715500100, 715513000, 715902000, 715907000

File Segment: EPI;

DWPI Class: T01

Manual Codes (EPI/S-X): T01-J05B

#### **Original Titles:**

...OVERVIEWS CORRESPONDING TO THE PLURALITY OF DOCUMENTS, RESOLVING

CO-REFERENTIALITY RELATED TO FREQUENCY WITHIN DOCUMENT, DETERMINING  
TOPIC STAMPS FOR EACH DOCUMENT SEGMENTS...

Original Publication Data by Authority

Original Abstracts:

...particular interest to the user. In a preferred embodiment, the capsule overviews include a containment **hierarchy** which relates the **different** information levels in a document together, and which includes a collection of highly salient topic stamps embedded in layers of **progressively** richer and more **informative** contextualized text fragments. The novel presentation metaphors which the invention utilizes are based on notions...

...particular interest to the user. In a preferred embodiment, the capsule overviews include a containment **hierarchy** which relates the different information levels in a document together, and which includes a collection of highly salient **topic** stamps embedded in layers of **progressively** richer and more informative contextualized **text** fragments. The **novel** presentation metaphors which the invention utilizes are based on **notions** of temporal typography, in particular for exploiting the interactions between form and content...

...particular interest to the user. In a preferred embodiment, the capsule overviews include a containment **hierarchy** which relates the different information levels in a document together, and which includes a collection of highly salient topic stamps embedded in layers of **progressively** richer and more informative contextualized text fragments. The novel presentation metaphors which the **invention** utilizes are based on notions of temporal typography, in particular for exploiting the interactions between form and content.

...particular interest to the user. In a preferred embodiment, the capsule overviews include a containment **hierarchy** which relates the different information levels in a document together, and which includes a collection of highly salient topic stamps embedded in layers of **progressively** richer and more informative contextualized text fragments. The novel presentation metaphors which the invention utilizes...

...notions of temporal typography, in particular for exploiting the interactions between form and content.

...

...particular interest to the user. In a preferred embodiment, the capsule overviews include a containment **hierarchy** which relates the different information levels in a document together, and which includes a collection of highly salient topic stamps embedded in layers of **progressively** richer and more informative contextualized text fragments. The novel presentation metaphors which the invention utilizes...

...between form and content.

A method...

...read it in more detail. In a preferred embodiment, the capsule overviews include a containment **hierarchy** which relates the different information levels in a document together, and which includes a collection of highly salient topic stamps embedded in layers of **progressively** richer and more informative contextualized text fragments. The novel presentation metaphors

which the invention utilizes...

...form and content.

A method for...

...read it in more detail. In a preferred embodiment, the capsule overviews include a containment **hierarchy** which relates the different information levels in a document together, and which includes a collection of highly salient topic stamps embedded in layers of **progressively** richer and more informative contextualized text fragments. The novel presentation metaphors which the invention utilizes

**Claims:**

...document segments;3) resolving co-referentiality among the discourse referents within, and across, the document **segments** , wherein the resolving **step** comprises linking the discourse referents by co-referentiality with each other to assess a frequency...

...prominence;4) calculating salience values for the discourse referents based upon the resolving step;5) **determining topic stamps for** the document **segments** based upon discourse **salience** values of the associated **discourse** referents; and6) providing a capsule overview of the document constructed from the topic stamps; andc) dynamically delivering document content as **encapsulated** within the plurality of capsule overviews...

... What is **claimed** is:1. A computer readable medium containing programming instructions for dynamically presenting the contents of...

...document segments;3) resolving co-referentiality among the discourse referents within, and across, the document **segments** , wherein the resolving **step** comprises linking the discourse referents by co-referentiality with each other to assess a frequency...

...4) calculating discourse salience values for the discourse referents based upon the resolving step;5) **determining topic stamps for** each of the document **segments** based upon discourse **salience** values of the associated discourse referents; and6) providing a capsule overview of the document,

10/69,K/11 (Item 11 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2007 The Thomson Corporation. All rts. reserv.

0009313166 - Drawing available  
WPI ACC NO: 1999-244130/199920  
Related WPI Acc No: 1999-045091  
XRPX Acc No: N1999-181663

**collection selection relative to a set of databases to obtain consistent relative-ranking collection selection results each iteration**

Patent Assignee: INFOSEEK CORP (INFO-N)

Inventor: CHANG W I; KIRSCH S T

**Patent Family** (4 patents, 79 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
WO 1999014691	A1	19990325	WO 1998US18844	A	19980910	199920 B
AU 199892282	A	19990405	AU 199892282	A	19980910	199933 E
US 5983216	A	19991109	US 1997928294	A	19970912	199954 E
US 6018733	A	20000125	US 1997928543	A	19970912	200012 E

Priority Applications (no., kind, date): (US 1997928543 A 19970912; US 1997928542 A 19970912; US 1997928294 A 19970912)

#### Patent Details

Number	Kind	Lan	Pg	Dwg	Filing	Notes
WO 1999014691	A1	EN	46	6		

National Designated States, Original: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH GM HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG UZ VN YU ZW

Regional Designated States, Original: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW

AU 199892282 A EN Based on OPI patent WO 1999014691

#### Alerting Abstract WO A1

NOVELTY - A collection selection query including a **set of set search terms** is obtained. An inverse collection frequency is determined for each search term with respect to each database and the set of databases. A document frequency is determined for each search term with respect to each database. A ranking value is determined for each database based on a sum of the products of the inverse collection frequencies for the search terms and the document frequencies for respective search terms. A subset of the set of databases is selected based on set criteria dependent on the ranking value for each database.

DESCRIPTION - The method involves: a) obtaining a collection selection query including a **set of set search terms**, b) determining an inverse collection frequency for each search term with respect to each database and the set of databases, and determining a document frequency for each search term with respect to each database, c) determining a ranking value for each database based on a sum of the products of the inverse collection frequencies for the search terms and the document frequencies for respective search terms, d) selecting a subset of the set of databases based on set criteria dependent on the ranking value for each database, and e) selectively repeating **portions** of the **steps** (b) through (d) with respect to each search term for each iteration of the method.

USE - The method is used to permit iterative performance of collection selection relative to a set of databases, where each database includes several documents, to obtain consistent relative-ranking collection selection results each iteration.

ADVANTAGE - Improves selection of most relevant collections for searching



based on an ad hoc query.

DESCRIPTION OF DRAWINGS - The drawing shows a flow diagram illustrating the operation in supporting a meta-index database construction and user search.

**Title Terms/Index Terms/Additional Words:** COLLECT; SELECT; RELATIVE; SET; OBTAIN; CONSISTENT; RANK; RESULT; ITERATIVE

#### **Class Codes**

International Classification (Main): G06F-017/30

US Classification, Issued: 707002000, 707003000, 707004000, 707005000, 707003000, 707001000, 707004000, 707005000, 707102000, 707103000

File Segment: EPI;

DWPI Class: T01; W01

Manual Codes (EPI/S-X): T01-H07C5E; T01-J05B; T01-J05B1; W01-A06B7

#### **Original Titles:**

...Methods for iteratively and interactively performing **collection** selection in full **text** searches...

...METHODS FOR ITERATIVELY AND INTERACTIVELY PERFORMING **COLLECTION** SELECTION IN FULL **TEXT** SEARCHES...

**Alerting Abstract ...NOVELTY** - A collection selection query including a **set** of **set** search **terms** is obtained. An inverse collection frequency is determined for each search term with respect to...

DESCRIPTION - The method involves: a) obtaining a collection selection query including a **set** of **set** search **terms**, b) determining an inverse collection frequency for each search term with respect to each database...

...on set criteria dependent on the ranking value for each database, and e) selectively repeating **portions** of the **steps** (b) through (d) with respect to each search term for each iteration of the method...

#### **Original Publication Data by Authority**

#### **Original Abstracts:**

...by exclusion of predetermined context-free single-word terms and punctuation; (b) applying each such **selected** term against a **meta - index** descriptive of the document collections; (c) determining cumulative rankings for the document collections relative to each **such** selected **term** normalized against the **plurality** of document **collections**; and (d) selecting a set of the document collections having the highest relative cumulative rankings...

...repetitive steps of determining an inverse collection frequency and a document frequency for each database; **determining** a ranking value for **each** database; **selecting** a **subset** of the set of databases based on predetermined criteria dependant on the ranking value for...

...file that describes the query significant search terms that are present in a particular document **collection** **correlated** to normalized document usage frequencies of such **terms** within the documents of each document **collection**. By **access** to the meta-information data file, a relevance score for each of the document collections...

...determined. The method then returns an identification of the subset of the plurality of document **collections** having the highest **relevance**

scores for use in evaluating the predetermined query. The meta-information data file may be constructed to include document normalized term frequencies and other contextual information that can be...

**Claims:**

...and information about documents in the corresponding ones of the document collections; parsing said input query text to select single-word terms and multiple-word phrase terms from said query text by...

...word terms and punctuation; applying each such selected term against the meta-index values in said meta-index to determine correlation between the selected terms and the meta-index values; determining cumulative rankings for said document collections based upon said correlation relative to each such selected term normalized against said plurality of document collections; and selecting a subset of said document collections having the highest relative cumulative rankings whereby said subset of...

...collections to search using said input query text, searching each of said subset of document collections with said input query text to select documents correlating to said query text.

...

...each iteration, said method comprising the steps of: a) obtaining a collection selection query including a set of predetermined search terms; b) determining an inverse collection frequency for each member of said set of predetermined search terms with respect to each said database and said set of databases, and determining a document frequency for each member of said set of predetermined search terms with respect to each said database; c) determining a ranking value for each said database based on a sum of the products of said inverse collection frequencies for said set of predetermined search terms and said document frequencies for respective members of said set of search terms; d) selecting a subset of said set of databases based on predetermined criteria dependant on said ranking value for each said database; and e) selectively repeating portions of said steps (b) through (d) with respect to each member of said set of predetermined search terms for each iteration of said method.

10/69,K/12 (Item 12 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2007 The Thomson Corporation. All rts. reserv.

0007518171 - Drawing available  
WPI ACC NO: 1996-131256/199614  
XRPX Acc No: N1996-110318

**Automatic recognition of language carrying numeric data - applies sequence of languages to part of data stream to look for characteristic patterns or signatures to decide translator module**

Patent Assignee: OCE GRAPHICS FRANCE SA (CHEZ); OCE NEDERLAND BV (CHEZ);  
OCE-TECHNOLOGIES BV (CHEZ)

Inventor: EVEN R; GENETIER L; LUC G; RENEKA E; ROBERTUS; TILLAART R C W T M  
V D; VAN DEN TILLAART R C W; VAN DEN TILLAART R C W T M

**Patent Family** (11 patents, 10 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
FR 2723457	A1	19960209	FR 19949801	A	19940808	199614 B
EP 702289	A1	19960320	EP 1995202139	A	19950804	199616 E
JP 8123636	A	19960517	JP 1995202282	A	19950808	199630 E
TW 311997	A	19970801	TW 1995107070	A	19950708	199745 E
CN 1125882	A	19960703	CN 1995109288	A	19950808	199748 E
US 5960113 <sup>1</sup>	A	19990928	US 1995507554	A	19950726	199947 E
JP 3195522	B2	20010806	JP 1995202282	A	19950808	200147 E
EP 702289	B1	20011114	EP 1995202139	A	19950804	200175 E
DE 69523848	E	20011220	DE 69523848	A	19950804	200207 E
			EP 1995202139	A	19950804	
KR 408762	B	20040522	KR 199524351	A	19950807	200461 E
CN 1160613	C	20040804	CN 1995109288	A	19950808	200612 E

Priority Applications (no., kind, date): FR 19949801 A 19940808

#### Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
FR 2723457	A1	FR	35		
EP 702289	A1	EN	22		
Regional Designated States, Original: DE FR GB IT NL					
JP 8123636	A	JA	21		
TW 311997	A	ZH			
JP 3195522	B2	JA	21		Previously issued patent JP 08123636
EP 702289	B1	EN			
Regional Designated States, Original: DE FR GB IT NL					
DE 69523848	E	DE			Application EP 1995202139
					Based on OPI patent EP 702289
KR 408762	B	KO			Previously issued patent KR 96008607

#### Alerting Abstract FR A1

The automatic recognition method selects a block of data from the incoming data stream. It searches the block for indications of the presence or absence of the particular language in the data block. The data block is tested for a variety of known languages, the languages being applied in a predetermined order.

The order of application of the languages is by increasing probability of an error of recognition occurring. The testing procedure looks for a characteristic element in the language, or a signature.

Alternatively key words or particular synchronisation characters are sought. The width of the observation window is varied to suit the language being tested. From these test an interpretation module is selected for the numeric data stream.

USE/ADVANTAGE - Printing or display of drawings from tracer data.

Accurate and fast recognition of language used for transfer of numeric data.

**Title Terms/Index Terms/Additional Words:** AUTOMATIC; RECOGNISE; LANGUAGE; CARRY; NUMERIC; DATA; APPLY; SEQUENCE; PART; STREAM; CHARACTERISTIC; PATTERN; SIGNATURE; DECIDE; TRANSLATION; MODULE; DRAWINGS; PRINTING; DISPLAY; TRACER

#### **Class Codes**

International Classification (Main): G06F-015/38, G06F-017/40, G06F-003/12  
(Additional/Secondary): B41J-029/38, G06F-013/00, G06F-003/14, G06F-009/45, G06K-009/62

International Classification (+ Attributes)

IPC + Level Value Position Status Version

G06F-0017/27 A I R 20060101

G06F-0003/12 A I R 20060101

G06F-0017/27 C I R 20060101

G06F-0003/12 C I R 20060101

US Classification, Issued: 382229000, 382181000, 395112000, 395114000, 395500000, 395707000

File Segment: EngPI; EPI;

DWPI Class: T01; P75; P86

Manual Codes (EPI/S-X): T01-C05A; T01-C05B; T01-D02; T01-J11; T01-S

...applies sequence of languages to part of data stream to look for characteristic patterns or signatures to decide translator module

#### **Original Publication Data by Authority**

#### **Original Abstracts:**

...special synchronization characters or keywords, and then for languages using mnemonics made up of a **determined** number of significant **characters**. The method **is used** for automatically selecting an interpreter module for decoding the received data, in particular the data and then for languages using mnemonics made up of a **determined** number of significant characters. The method is used for automatically selecting an interpreter module for...

#### **Claims:**

...in that</b> recognition is performed by searching for a plurality of known languages in a **order of** increasing probability of recognition error proceeding, for each language, with a search in the data...block that tend to indicate the presence or the absence of a language;said step of seeking including the **sub - steps** ofsequentially testing for a plurality of known languages according to a predetermined sequential arrangement...

...andproceeding, for each one of said known languages, with a search in said block **for at least one language element characteristic** of said one of said known languages.

13/69,K/1 (Item 1 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2007 The Thomson Corporation. All rts. reserv.

0014143537 - Drawing available  
WPI ACC NO: 2004-328314/200430  
Related WPI Acc No: 2004-601774; 2007-506612  
XRPX Acc No: N2004-261954

**Hypertext link installation system for Internet, has link installation unit to insert hypertext links associated with occurrences into document and output unit to return document corresponding to document with inserted links**

Patent Assignee: SMITH J O (SMIT-I)

Inventor: SMITH J O

**Patent Family** (2 patents, 1 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
US 20040068527	A1	20040408	US 1998103089	P	19981005	200430 B
			US 1999412248	A	19991005	
			US 2003681621	A	20031007	
US 7233950	B2	20070619	US 1998103089	P	19981005	200741 E
			US 1999412248	A	19991005	
			US 2003681621	A	20031007	

Priority Applications (no., kind, date): ~~US 1998103089~~ P ~~19981005~~; US 1999412248 A 19991005; US 2003681621 A 20031007

#### Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
US 20040068527	A1	EN	53	8	Related to Provisional US 1998103089 Division of application US 1999412248
US 7233950	B2	EN			Related to Provisional US 1998103089 Division of application US 1999412248  Division of patent US 6772139

#### Alerting Abstract US A1

NOVELTY - The system has a string matching unit to find occurrences in a document of **key phrases** associated with a **selected** subset of hypertext links (101, 102, 103, 104, 105). A link installation unit inserts the hypertext links associated with the occurrences into the document submitted for hypertext link installation. An output unit returns a document corresponding to the document including the inserted hypertext links.

DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

1.a method for installing hypertext links in a document

2.a system for retrieving hypertext links.

USE - Used for installing a hypertext link in a document for access on the Internet through World Wide Web.

ADVANTAGE - The link installation unit provides the link installation service, thereby automatically installing hypertext links within information submitted to the service by the hypertext authors.

DESCRIPTION OF DRAWINGS - The drawing shows an example initial web page seen by a visitor using a web browser to access the online version of the service.

101, 102, 103, 104, 105Hypertext links

**Title Terms/Index Terms/Additional Words:** LINK; INSTALLATION; SYSTEM; UNIT;  
INSERT; ASSOCIATE; OCCUR; DOCUMENT; OUTPUT; RETURN; CORRESPOND

**Class Codes**

International Classification (+ Attributes)

IPC + Level Value Position Status Version

G06F-0017/00 A I L B 20060101

G06F-0017/30 A I R 20060101

G06F-0007/00 A I F B 20060101

G06F-0017/00 C I B 20060101

G06F-0017/30 C I R 20060101

G06F-0007/00 C I B 20060101

US Classification, Issued: 707204000, 707010000, 707104100

File Segment: EPI;

DWPI Class: T01

Manual Codes (EPI/S-X): T01-N03B2A

**Alerting Abstract** ...NOVELTY - The system has a string matching unit to find occurrences in a document of **key phrases** associated with a **selected** subset of hypertext links (101, 102, 103, 104, 105). A link installation unit inserts the...  
...a method for installing hypertext links in a documenta **system** for **retrieving** hypertext links...

**Original Publication Data by Authority**

**Claims:**

...is claimed is:<b>1</b>. A system for installing hypertext links in a document, comprisinga) **hierarchical** database means containing hypertext links, wherein each of said hypertext links is associated with a **set** of **key phrases** ;b) link **selection** means for **selecting** a subset of said hypertext links, thereby forming a selected subset of hypertext links;c...

...hypertext link installation;d) string matching means for finding occurrences in said submitted document of **key phrases** associated with said **selected** subset of hypertext links;e) link installation means for inserting into said submitted document hypertext...

...a submitter;matching means for finding an occurrence of at least one of the stored **key phrases** in the submitted **text** , thereby **determining** a **set** of one or more found key phrases;link retrieval means for retrieving from the database...

13/69,K/3 (Item 3 from file: 350)

DIALOG(R)File 350:Derwent WPIX

(c) 2007 The Thomson Corporation. All rts. reserv.

0013087019 - Drawing available

WPI ACC NO: 2003-167749/200316

XRPX Acc No: N2003-132497

**Internet high-speed access search engines for information retrieval systems where the query processor performs the step of analyzing using a hybrid method based on linguistic and mathematical approaches for an automatic text categorization**

Patent Assignee: COGISUM INTERMEDIA AG (COGI-N); MEIK F (MEIK-I); WIELSCH M (WIEL-I)

Inventor: MEIK F; WIELSCH M

**Patent Family** (7 patents, 95 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
WO 2003005235	A1	20030116	WO 2001EP7649	A	20010704	200316 B
EP 1402408	A1	20040331	EP 2001967123	A	20010704	200424 E
			WO 2001EP7649	A	20010704	
KR 2004013097	A	20040211	WO 2001EP7649	A	20010704	200438 E
			KR 2004700048	A	20040102	
AU 2001287582	A1	20030121	AU 2001287582	A	20010704	200452 E
			WO 2001EP7649	A	20010704	
JP 2004534324	W	20041111	WO 2001EP7649	A	20010704	200474 E
			JP 2003511133	A	20010704	
CN 1535433	A	20041006	CN 2001823447	A	20010704	200507 E
			WO 2001EP7649	A	20010704	
US 20050108200	A1	20050519	WO 2001EP7649	A	20010704	200534 E
			US 2004482833	A	20041220	

Priority Applications (no., kind, date): WO 2001EP7649 A 20010704

#### Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
--------	------	-----	----	-----	--------------

WO 2003005235	A1	EN.	121	14	
---------------	----	-----	-----	----	--

National Designated States,Original: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW

Regional Designated States,Original: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW

EP 1402408	A1	EN			PCT Application WO 2001EP7649
					Based on OPI patent WO 2003005235

Regional Designated States,Original: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI TR

KR 2004013097	A	KO			PCT Application WO 2001EP7649
AU 2001287582	A1	EN			PCT Application WO 2001EP7649
					Based on OPI patent WO 2003005235
JP 2004534324	W	JA	191		PCT Application WO 2001EP7649
					Based on OPI patent WO 2003005235
CN 1535433	A	ZH			PCT Application WO 2001EP7649
US 20050108200	A1	EN			PCT Application WO 2001EP7649

#### Alerting Abstract WO A1

NOVELTY - An interactive document **retrieval system** where the query processor performs the step of analyzing using a hybrid method based on linguistic and mathematical approaches for an automatic text categorization. The captured document is analyzed to determine their **text**

**patterns** which are commonly occurring and searchable phrases, pairing of words, with each pairing comprising two searchable words, where one word in each pairing occurs frequently within the document and the other word in each pairing occurs near the one word frequently within the document.

DESCRIPTION - The knowledge base is initially constructed by analyzing indexed documents to which **topics** have previously been assigned, thereby **determining** the indexed documents **word patterns**, and then storing in the knowledge database these **word patterns** for the indexed documents and the topics assigned to these documents, and then relating the **word pattern** of an indexed document to the topic assigned to that same indexed document.

INDEPENDENT CLAIMS are also included for the following:

1. An interactive method of searching for and retrieving documents after receiving a search query from a requestor.
2. A computer program.
3. A mobile computing and/or telecommunications device comprising a graphical user interface capable of applying the WAP standard for accessing documents from the Internet and/or any corporate network.

USE - High-speed access search engines for information **retrieval systems** used in the Internet and/or **corporate intranet** domains for **retrieving** accessible documents using automatic text categorization techniques to support the presentation of search query results within high-speed network environments.

ADVANTAGE - Provides an integrated, automatic and open information **retrieval system**, comprising a hybrid method based on **linguistic and mathematical** approach for an automatic text categorization.

Enables the possibility of meeting the requirements of all Internet users by means of the novel Internet archive. Provides desired information in a quick, simple and accurate manner that allows significant advantages with regard to data management within individual companies.

DESCRIPTION OF DRAWINGS - The drawing is an overview block diagram of an indexed extensible, interactive **retrieval system**.

**Title Terms/Index Terms/Additional Words:** HIGH; SPEED; ACCESS; SEARCH; ENGINE; INFORMATION; RETRIEVAL; SYSTEM; QUERY; PROCESSOR; PERFORMANCE; STEP; HYBRID; METHOD; BASED; MATHEMATICAL; APPROACH; AUTOMATIC; TEXT

#### **Class Codes**

International Classification (Main): G06F-017/30

US Classification, Issued: 707003000

File Segment: EPI;

DWPI Class: T01; W01

Manual Codes (EPI/S-X): T01-J05B1; T01-J05B3; T01-J05B4P; T01-J16A;

T01-N03A2; T01-S03; W01-C01G6E

**Internet high-speed access search engines for information retrieval systems where the query processor performs the step of analyzing using a hybrid method based on...**

#### **Original Titles:**

CATEGORY BASED, EXTENSIBLE AND INTERACTIVE **SYSTEM** FOR DOCUMENT **RETRIEVAL**  
...

...CATEGORY BASED, EXTENSIBLE AND INTERACTIVE **SYSTEM** FOR DOCUMENT **RETRIEVAL**



...

...Category based, extensible and interactive **system** for document **retrieval**

...

...CATEGORY BASED, EXTENSIBLE AND INTERACTIVE **SYSTEM** FOR DOCUMENT **RETRIEVAL**

**Alerting Abstract** ...NOVELTY - An interactive document **retrieval system** where the query processor performs the step of analyzing using a hybrid method based on...

...mathematical approaches for an automatic text categorization. The captured document is analyzed to determine their **text patterns** which are commonly occurring and searchable phrases, pairing of words, with each pairing comprising two...

DESCRIPTION - The knowledge base is initially constructed by analyzing indexed documents to which **topics** have previously been assigned, thereby **determining** the indexed documents **word patterns**, and then storing in the knowledge database these **word patterns** for the indexed documents and the topics assigned to these documents, and then relating the **word pattern** of an indexed document to the topic assigned to that same indexed document...

...USE - High-speed access search engines for information **retrieval systems** used in the Internet and/or **corporate intranet** domains for **retrieving** accessible documents using automatic text categorization techniques to support the presentation of search query results...

...ADVANTAGE - Provides an integrated, automatic and open information **retrieval system**, comprising a hybrid method based on **linguistic** and mathematical approach for an automatic text categorization...

...DESCRIPTION OF DRAWINGS - The drawing is an overview block diagram of an indexed extensible, interactive **retrieval system**.

#### Original Publication Data by Authority

##### Original Abstracts:

An integrated, automatic and open information **retrieval system** (100) comprises an **hybrid** method based on linguistic and mathematical approaches for an automatic text categorization. It solves the problems of conventional systems by combining an automatic content recognition technique with a self-learning **hierarchical** scheme of indexed **categories**. In response to a word submitted by a requestor, said **system** (100) **retrieves** documents containing that **word**, analyzes the documents to determine their **word-pair patterns**, matches the document **patterns** to database patterns that are related to topics, and thereby assigns topics to each document...

...assigned to more than one topic, a list of the document topics is presented to the requestor, and the requestor **designates** the relevant **topics**. The requestor is **then** granted access only to documents assigned to relevant topics. A knowledge database (1408) linking search...

...In information **retrieval** (IR) **systems** with high-speed access, especially to **search engines** **applied** to the Internet and/or corporate intranet domains for retrieving accessible documents automatic text categorization...

...of search query results within high-speed network environments. An integrated, automatic and open information **retrieval system** (<b>100</b>) comprises an hybrid method based on linguistic and mathematical approaches for an **automatic** text categorization. It solves the problems of conventional systems by combining an automatic content recognition technique with a self-learning **hierarchical** scheme of indexed categories. In response to a word submitted by a requester, said **system0** (<b>100</b>) **retrieves** documents containing that word, analyzes the documents to determine their **word-pair patterns**, **matches** the document patterns to database **patterns** that are related to **topics**, and thereby assigns **topics** to **each** document. If the retrieved documents are assigned to more than one topic, a list of the document topics is presented to the requester, and the requester **designates** the relevant **topics**. The requester is then granted access only to documents assigned to relevant **topics**. A **knowledge** database (<b>1408</b>) **linking** search terms to documents and documents to topics is established and maintained to speed future...

...An integrated, automatic and open information **retrieval system** (100) comprises an hybrid method based on linguistic and mathematical approaches for an automatic text categorization. It solves the problems of conventional **systems** by **combining** an automatic content recognition technique with a self-learning **hierarchical** scheme of indexed categories. In response to a word submitted by a **requestor**, said **system** (100) **retrieves** documents containing that word, analyzes the documents to determine their word-pair **patterns**, matches the document **patterns** to database **patterns** that are **related** to **topics**, and thereby assigns **topics** to each document. If the retrieved documents are assigned to more than one topic, a list of the document topics is presented to the **requestor**, and the requestor **designates** the relevant **topics**. The requestor is then granted access only to documents assigned to relevant topics. A knowledge database (1408) linking search terms to **documents** and documents to **topics** is established and maintained to speed future searches. Additionally, new strategies are presented to deal...

...de la combinaison d'une technique de reconnaissance de contenu classique et d'un processus **hierarchique** d'auto-apprentissage de categories indexees. En reponse a un mot propose par un demandeur, le systeme (100) recupere des documents contenant ce meme mot, analyse les documents pour **determiner** leurs structures de paires de mots, etablit une correspondance entre les structures de documents et...

**Claims:**

<b>1</b>. An interactive document **retrieval system** (<b>100</b>) designed to **search** for documents after receiving a search query from a requestor, said system comprising: a knowledge...

...least one data structure (<b>202</b>, <b>208</b>, <b>210</b>, <b>212</b>, <b>214</b>, <b>216</b> and/or <b>218</b>) that relates **text patterns** to topics, and a query processor (<b>400</b>) that, in response to the receipt of a search query from a requester, **performs** the following steps: searching for and trying to capture documents containing at least one term related...

...the search query, if any documents are captured, analyzing the captured documents to determine their **text patterns**, categorizing the captured documents by comparing **each** document's **text pattern** to the **text patterns** in the knowledge database (<b>200</b>), and if a document's **text pattern** is similar to a **text pattern** in the knowledge database (<b>200</b>), assigning to that document the **similar word pattern**'s

related topic, presenting at least one list of the topics assigned to the categorized documents to the requester, and asking the requester to designate at least one topic from the list as a topic that is relevant to the requestor's search, and granting the requestor access to the subset of captured and categorized documents to which topics designated by the requestor have been assigned, wherein the word patterns determined by analysis are pairings of words, each pairing comprising two searchable words with one word occurring frequently within the document and the other word occurring near the one word frequently within the document.

13/69,K/4 (Item 4 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2007 The Thomson Corporation. All rts. reserv.

0010791617 - Drawing available  
WPI ACC NO: 2001-407058/200143  
XRPX Acc No: N2001-301085

**Computer implemented document topic arrangement for information retrieval system , involves displaying set of topics having semantic correspondence with topic selected initially, along with preset parameters**

Patent Assignee: HOROWITZ D (HORO-I); SHEFNER D B (SHEF-I)

Inventor: HOROWITZ D; SHEFNER D B

**Patent Family** (1 patents, 1 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
US 6236987	B1	20010522	US 199854753	A	19980403	200143 B

Priority Applications (no., kind, date): US-199854753 A 19980403

#### Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
US 6236987	B1	EN	30	12	

#### Alerting Abstract US B1

NOVELTY - A set of documents (152) which satisfies the user's query, is received by user interface (110) such as keyboard and a **topic** related to user query is **selected** from received document. According to user **selection** of **topic** arrangements, the **topics** having semantic correspondence to **selected topic** are **selected** , arranged and displayed along with preset parameters.

DESCRIPTION - An INDEPENDENT CLAIM is also included for the information **retrieval system** .

USE - Used for information **retrieval system** in information domains such as document management systems, library catalog, search engines for the world wide web, database, etc.

ADVANTAGE - Improves understanding of organization, relationships and nature of content in a document collection through distinct topic arrangements, according to user interested queries and for interactively constructing topic and key word based queries for further navigating the document collection, is achieved.

DESCRIPTION OF DRAWINGS - The figure shows the operation of the information **retrieval system** .

110 User interface

152 Documents

**Title Terms/Index Terms/Additional Words:** COMPUTER; IMPLEMENT; DOCUMENT; TOPIC; ARRANGE; INFORMATION; RETRIEVAL; SYSTEM; DISPLAY; SET; CORRESPOND; SELECT; INITIAL; PRESET; PARAMETER

#### Class Codes

International Classification (Main): G06F-017/30

US Classification, Issued: 707003000, 707002000, 707004000, 707005000, 705035000, 705008000, 706045000

File Segment: EPI;

DWPI Class: T01

Manual Codes (EPI/S-X): T01-H07C5E; T01-J05B2; T01-J05B3; T01-J05B4P

**Computer implemented document topic arrangement for information retrieval**

system , involves displaying set of topics having semantic correspondence with topic selected initially, along with preset parameters

**Original Titles:**

Dynamic content organization in information retrieval systems .

**Alerting Abstract** ...the user's query, is received by user interface (110) such as keyboard and a **topic** related to user query is **selected** from received document. According to user **selection** of **topic** arrangements, the **topics** having semantic correspondence to **selected topic** are **selected** , arranged and displayed along with preset parameters. DESCRIPTION - An INDEPENDENT CLAIM is also included for the information retrieval system .

...

...USE - Used for information **retrieval system** in information domains such as document management systems, library catalog, search engines for the world...

...DESCRIPTION OF DRAWINGS - The figure shows the operation of the information **retrieval system** .

**Original Publication Data by Authority**

**Original Abstracts:**

...plurality of topics. Each topic expresses an idea or concept, and is associated with a **set** of **terms** which describe the topic, a set of documents in the document collection which are about the topic. Each topic also has **topic** -subtopic relationships with **selected** other **topics** , forming local **topic hierarchies** . A query analysis module receives a current query and processes the query against the document...

...module processes the document set according to defined parameters and a user selection or automatic **selection** of a desired **topic** arrangement to create various types of topic arrangements. These topic arrangements include supertopics, subtopics, perspective...

**Claims:**

...comprising:processing the query to select a set of documents satisfying the query;receiving a **selection** of at least one **topic derived** from the query; **determining** the supertopic arrangement as a combination of supertopics that are associated with the documents of the document set and with the **selected topic** and that optimally generalizes the document **set with** respect to parameters; anddisplaying the supertopic arrangement.

13/69,K/5 (Item 5 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2007 The Thomson Corporation. All rts. reserv.

0009579613 - Drawing available  
WPI ACC NO: 1999-527164/199944  
XRPX Acc No: N1999-390494

**Query processing method in multidocument search and retrieval system  
for books, articles, periodicals**

Patent Assignee: ORACLE CORP (ORAC-N)

Inventor: WICAL K

**Patent Family** (1 patents, 1 countries)

Patent			Application					
Number	Kind	Date	Number	Kind	Date	Update		
US 5953718	A	19990914	US 1997967774	A	19971112	199944	B	

Priority Applications (no., kind, date): US 1997967774 A 19971112 }

**Patent Details**

Number	Kind	Lan	Pg	Dwg	Filing Notes
US 5953718	A	EN	31	11	

**Alerting Abstract US A**

NOVELTY - Point of view gists are generated for atleast one document. A query is processed, which has a query term identifying topics related to the query. Point of view gists are selected from one or more documents in response to the query, to generate a new research document.

DESCRIPTION - Themes which define an overall content for the document are stored. The themes relevant to the query are selected as queries and documents that contain the themes are selected. An INDEPENDENT CLAIM is also included for the computer readable medium.

USE - In multidocument search and **retrieval system** for books, articles, periodicals.

ADVANTAGE - Emulates the paradigm of a researcher by extracting portions of different documents to infer an answer to the search query. Utilizes rich and comprehensive content processing system to accurately identify themes that define the content of the source material.

DESCRIPTION OF DRAWINGS - The figure shows the search and **retrieval system**.

**Title Terms/Index Terms/Additional Words:** QUERY; PROCESS; METHOD; SEARCH; RETRIEVAL; SYSTEM; BOOK; ARTICLE; PERIOD

**Class Codes**

International Classification (Main): G06F-017/30  
US Classification, Issued: 707005000, 707001000, 707002000, 707003000, 707004000, 707006000, 707007000, 707101000, 707104000

File Segment: EPI;

DWPI Class: T01

Manual Codes (EPI/S-X): T01-J05B

**Query processing method in multidocument search and retrieval system  
for books, articles, periodicals**

**Original Titles:**

Research mode for a knowledge base search and **retrieval system**.

**Alerting Abstract ...USE** - In multidocument search and **retrieval system** for books, articles, periodicals...

...DESCRIPTION OF DRAWINGS - The figure shows the search and **retrieval system** .

**Original Publication Data by Authority**

**Original Abstracts:**

A research mode in a search and **retrieval system** generates a **research document** that infers an answer to a query from multiple documents. The search and **retrieval system** includes point of view gists for documents to provide a synopsis for a corresponding document with a slant toward a topic. To generate a research document, the search and **retrieval system** processes a query to identify one or more **topics** related to the query, selects document themes relevant to the query, and then selects point of view gists, based on the document themes, that...

...slant towards the topics related to the query. A knowledge base, which includes categories arranged **hierarchically**, is configured as a directed graph to links those categories having a lexical, semantic or usage association. Through use of the knowledge base, an expanded **set** of query **terms** are generated, and **research** documents are compiled that include point of view gists relevant to the expanded **set** of query **terms**. A content processing **system**, which identifies the themes for a document and classifies the document themes in categories of...

**Claims:**

A method for processing a query in a search and **retrieval system**, said **method** comprising the steps of:generating a plurality of point of view gists for at least...

...processing a query, which includes at least one query term, to identify a plurality of **topics** related to said query; and **selecting** a plurality of point of view gists from one or more documents to generate, in...

...research document, wherein said point of view gists selected comprise synopses with slants toward said **topics** related to said query.

13/69,K/7 (Item 7 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2007 The Thomson Corporation. All rts. reserv.

0009226763 - Drawing available  
WPI ACC NO: 1999-153249/199913  
XRPX Acc No: N1999-110522

**Information retrieval support system for RDBMS**

Patent Assignee: MATSUSHITA ELECTRIC IND CO LTD (MATU)  
Inventor: FUJITA T; KIKUCHI C; KINOSHITA T; OYAMA T; SHINOKI H  
**Patent Family** (1 patents, 1 countries)

Patent			Application			
Number	Kind	Date	Number	Kind	Date	Update
US 5870750	A	19990209	US 1997865309	A	19970529	199913 B

Priority Applications (no., kind, date): [US 1997865309 A 19970529]

**Patent Details**

Number	Kind	Lan	Pg	Dwg	Filing Notes
US 5870750	A	EN	46	28	

**Alerting Abstract US A**

NOVELTY - A **character chain pattern** containing code added data and information corresponding to each pattern are generated by pattern and information generators (15,17), respectively. A **hierarchical** division code is added to each keyword output from keyword generator to produce a keyword **character chain pattern**.

DESCRIPTION - A division code is added to each segment of data stored in memory (11) according to their **hierarchy** level. A specific number is allocated to each segment of data by an allocation unit (14). The usage frequency of each data segment corresponding to the set **hierarchical** division code is computed by a calculator (25). A series of keyword **character chain** 1patterns is obtained by arranging the primary and secondary patterns according to their division code and code added **keywords**. During retrieval, a **specific** character chain data corresponding to **keyword** chain pattern is collated by a collation unit according to preferential order set by setting unit to extract a series of particular character chain data. An index file representing character chain data with respect to chain patterns is produced by a production unit. Based on the input keyword, the keyword included in each character chain data is compared and corresponding record is retrieved and displayed on a display unit (23).

USE - For RDBMS.

ADVANTAGE - Each segment of character data is easily identified, by judging the division code of each **pattern** accordingly. A particular **character** chain data series corresponding to keyword patterns is generated at high speed by using set division code and number data. The data corresponding to keyword is also identified easily, during retrieval. Retrieval speed is raised, by reducing number of **character** chain **patterns** and **character** chain information in index file.

DESCRIPTION OF DRAWINGS - The figure shows the block diagram of information **retrieval support system**.

- 11 Memory
- 14 Allocation unit
- 15,17 Pattern and information generators
- 23 Display unit
- 25 Calculator

**Title Terms/Index Terms/Additional Words:** INFORMATION; RETRIEVAL; SUPPORT; SYSTEM



## Class Codes

International Classification (Main): G06F-017/30

US Classification, Issued: 707101000, 707100000, 707102000, 707001000,  
707006000, 707007000

File Segment: EPI;

DWPI Class: T01

Manual Codes (EPI/S-X): T01-J05B1; T01-J05B3; T01-J05B4B; T01-J05B4M;  
T01-J05B4P

**Information retrieval support system for RDBMS**

## Original Titles:

Information **retrieval** **system** for **retrieving** a record of data  
including a keyword.

**Alerting Abstract ...NOVELTY** - A **character** chain **pattern** containing code added data and information corresponding to each pattern are generated by pattern and information generators (15,17), respectively. A **hierarchical** division code is added to each keyword output from keyword generator to produce a keyword **character** chain **pattern** . ...code is added to each segment of data stored in memory (11) according to their **hierarchy** level. A specific number is allocated to each segment of data by an allocation unit (14). The usage frequency of each data segment corresponding to the set **hierarchical** division code is computed by a calculator (25). A series of keyword **character** chain **patterns** is obtained by arranging the primary and secondary patterns according to their division code and code added **keywords** . During retrieval, a **specific** character chain data corresponding to **keyword** chain pattern is collated by a collation unit according to preferential order set by setting...  
...Each segment of character data is easily identified, by judging the division code of each **pattern** accordingly. A particular **character** chain data series corresponding to keyword patterns is generated at high speed by using set...

...keyword is also identified easily, during retrieval. Retrieval speed is raised, by reducing number of **character** chain **patterns** and **character** chain information in index file...

...DESCRIPTION OF DRAWINGS - The figure shows the block diagram of information **retrieval** support **system** .

## Original Publication Data by Authority

### Original Abstracts:

...of adjacent data, so that a piece of code-added data is produced. Thereafter, a **character** chain **pattern** (C1, C2 ) of **each** pair of adjacent **characters** C1 and C2 in the code-added data is prepared, an occurrence frequency Fi of...  
...data H is calculated, and character chain information (F1,F2,H,L) corresponding to each **character** chain **pattern** (C1, C2 ) is **prepared** . Therefore, a index file in which pieces of character chain information prepared for the same...

...ends of the keyword to produce a code-added keyword, and a series of keyword **character** chain **patterns** is **prepared** from **the** code-added keyword in the same manner. Thereafter, a series of particular character

chain information (F1,F2,H,L) corresponding to the series of keyword **character chain patterns** on **condition** that **F2** of first information equals to **F1** of second information following the first information and data ...

**Claims:**

An information **retrieval system**, **comprising**: **data** record storing means for storing a plurality of data records, pieces of data respectively composed...

...a top position of the piece of data and a current position of the current **character**; **character chain pattern** preparing means for preparing a first **character chain pattern** of each **pair** of **characters** adjacent to each other in one **code**-added data produced by the division code adding means for each code-added data and preparing a second **character chain pattern** of each pair of one division code and one **character** adjacent to the division code in one code-added data for each code-added data; character chain...

...preparing a piece of character chain information corresponding to each of the first and second **character chain patterns** prepared by the **character chain pattern** preparing means according to the **occurrence frequencies** obtained by the **character occurrence frequency calculating** means and the data numbers and the record numbers allocated by the number allocating means...

...one character and one division code of one piece of data corresponding to one first or second **character chain pattern**, one data number of the piece of data and one record **number** of the piece of data; index file preparing means for preparing an index file for the particular...

...the character chain information preparing means are listed in correspondence to the first and second **character chain patterns** prepared by the **character chain pattern** preparing means; keyword preparing means for preparing a keyword **composed** of a plurality of **characters**; **keyword** division code adding means for adding the division code to one end or both ends...

...prepared by the keyword preparing means to produce a piece of code-added keyword; keyword **character chain pattern** preparing means for preparing a first keyword **character chain pattern** of each pair of characters adjacent to each other in the **code**-added keyword produced by the keyword division **code** adding means, preparing a second keyword **character chain pattern** of each pair of one division code and one character adjacent to the division code in the code-added **keyword**, and preparing a series of keyword **character chain patterns** by arranging the first and second keyword **character chain patterns** in the **order** of the **characters** and division codes in the **code**-added keyword; data record retrieving means for **extracting** a **series** of particular **character chain information** corresponding to the series of keyword **character chain patterns** prepared by the keyword **character chain pattern** preparing means from the index file for the particular item prepared by the index **file** preparing means on condition that a **plurality** of **data** numbers in the pieces of particular character chain information agree with each other, a plurality...

13/69,K/8 (Item 8 from file: 350)  
 DIALOG(R)File 350:Derwent WPIX  
 (c) 2007 The Thomson Corporation. All rts. reserv.

0008712343 - Drawing available  
 WPI ACC NO: 1998-253111/199823  
 Related WPI Acc No: 1998-068087; 1998-221955  
 XRPX Acc No: N1998-199934

**Pattern recognition apparatus especially for use in pen input PC - has word dictionary which is searched for word defined by identification information which corresponds to optimised candidate character string**

Patent Assignee: GUNJI K (GUNJ-I); HITACHI LTD (HITA); KATSURA K (KATS-I); KUZUNUKI S (KUZU-I); MIURA M (MIUR-I); YOKOTA T (YOKO-I)  
 Inventor: GUNJI K; KATSURA A; KATSURA K; KEIKO G; KOYO K; KUZUNUKI S; MIURA M; SOSHIRO K; YOKOTA T

**Patent Family** (14 patents, 9 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
EP 841630	A2	19980513	EP 1997108132	A	19970520	199823 B
JP 10105571	A	19980424	JP 1996261936	A	19961002	199827 E
KR 1997076347	A	19971212	KR 199718940	A	19970516	199849 E
US 6097841	A	20000801	US 1997859410	A	19970520	200039 E
TW 421764	A	20010211	TW 1997114382	A	19970429	200146 E
JP 3205516	B2	20010904	JP 1996261936	A	19961002	200152 E
US 20010028742	A1	20011011	US 1997859410	A	19970520	200162 E
			US 2001789820	A	20010222	
CN 1173684	A	19980218	CN 1997113595	A	19970521	200170 E
TW 490643	A	20020611	TW 1997105652	A	19970429	200321 E
EP 841630	B1	20030917	EP 1997108132	A	19970520	200369 E
DE 69724910	E	20031023	DE 69724910	A	19970520	200377 E
			EP 1997108132	A	19970520	
US 6751605	B2	20040615	US 1997859410	A	19970520	200439 NCE
			US 2001789820	A	20010222	
CN 1143240	C	20040324	CN 1997113595	A	19970521	200609 E
KR 475266	B	20050711	KR 199718940	A	19970516	200660 E

Priority Applications (no., kind, date): JP 1996125360 A 19960521; JP 1996224808 A 19960827; JP 1996261936 A 19961002; EP 1997108132 A 19970520; US 2001789820 A 20010222

#### Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
EP 841630	A2	EN	98	47	
Regional Designated States,Original: DE FR GB NL					
JP 10105571	A	JA	12	9	
TW 421764	A	ZH			
JP 3205516	B2	JA	11		Previously issued patent JP 10105571
US 20010028742	A1	EN			Division of application US 1997859410
TW 490643	A	ZH			
EP 841630	B1	EN			
Regional Designated States,Original: DE FR GB NL					
DE 69724910	E	DE			Application EP 1997108132
					Based on OPI patent EP 841630
US 6751605	B2	EN			Division of application US 1997859410
					Division of patent US 6097841
KR 475266	B	KO			Previously issued patent KR 97076347

#### Alerting Abstract EP A2

The character recognition device includes a word dictionary (a6) which stores word identification information and **hierarchy** information for layering words into a **hierarchy** and for recognising each of the words within the **hierarchy**. A character transition probability table (a4) stores at least probabilities of transitions from any one character to another, and those pieces of the word identification information which correspond to **combinations** of **characters** resulting from the transitions.

A character transition probability table (a4) is used in optimising candidate **character strings** obtained by the **character** recognition device. The word dictionary is searched for words defined by those pieces of the word identification information which correspond to the optimised candidate string, thereby retrieving the searched words which are identified by the applicable pieces of the **hierarchy** information and which have yet to be input.

USE - Processing slips, invoices, forms etc.

ADVANTAGE - Eliminates need for operator to write whole of **character string** by hand. Fast operation. Allows lower **hierarchy** level data to be selected without knowing higher levels.

**Title Terms/Index Terms/Additional Words:** PATTERN; RECOGNISE; APPARATUS; PEN; INPUT; WORD; DICTIONARY; SEARCH; DEFINE; IDENTIFY; INFORMATION; CORRESPOND; OPTIMUM; CANDIDATE; CHARACTER; STRING

#### Class Codes

International Classification (Main): G06F-017/30, G06K-009/72

International Classification (+ Attributes)

IPC + Level Value Position Status Version

G06K-0009/72 A I R 20060101

G06K-0009/72 C I R 20060101

US Classification, Issued: 382229000, 382229000, 707003000, 707001000

File Segment: EPI;

DWPI Class: S05; T01; T04

Manual Codes (EPI/S-X): T01-J11A; T04-D07E; T04-F04

...dictionary which is searched for word defined by identification information which corresponds to optimised candidate character string

#### Original Titles:

...Apparatus for recognizing input **character strings** by inference...

...Apparatus for recognizing input **character strings** by inference...

... RETRIEVAL SYSTEM

...

...Apparatus for recognizing input **character strings** by inference...

...Apparatus for recognizing input **character strings** by inference...

...Apparatus for recognizing input **character strings** by inference

**Alerting Abstract** ...The character recognition device includes a word dictionary (a6) which stores word identification information and **hierarchy** information for layering words into a **hierarchy** and for recognising each of the words within the **hierarchy**. A character transition probability table (a4) stores at least probabilities of transitions from any one character to another, and those pieces of the word identification information which correspond to **combinations** of **characters** resulting

from the transitions...

...A character transition probability table (a4) is used in optimising candidate **character strings** obtained by the **character** recognition device. The word dictionary is searched for words defined by those pieces of the...

...string, thereby retrieving the searched words which are identified by the applicable pieces of the **hierarchy** information and which have yet to be input...

...ADVANTAGE - Eliminates need for operator to write whole of **character string** by hand. Fast operation. Allows lower **hierarchy** level data to be selected without knowing higher levels.

#### Original Publication Data by Authority

##### Original Abstracts:

An object of the present invention is to provide a character recognition apparatus for **inferring** the entire **character string** solely from a **user** -input handwritten keyword and displaying the inferred result as a candidate **character string** . The apparatus of the invention comprises: a **word** dictionary a6 storing word identification information and **hierarchy** information for layering a plurality of words into a **hierarchy** and for recognizing **each** of the words within the **hierarchy** ; a character transition **probability** table a4 storing probabilities of transitions from any one character to another, and those pieces of the word identification information which **correspond** to **combinations** of **characters** resulting from the **transitions** ; and an optimization unit for using the character transition probability table a4 in optimizing candidate **character strings** obtained by a **recognition** unit. The **word** dictionary a6 is searched for a word defined by the word identification information which corresponds to the optimized candidate **character string** , whereby the **searched word** is retrieved which **applies** to the **hierarchy** information and which **has** yet to be input...

...of the present invention is to provide a character recognition apparatus for inferring the entire **character string** solely from a user-input handwritten keyword and displaying the inferred result as a candidate **character string** . The apparatus of the invention comprises: a **word dictionary** a<b>6 </b>storing **word** identification information and **hierarchy** information for layering a plurality of words into a **hierarchy** and for recognizing **each** of the **words** within the **hierarchy** ; a character transition probability table a<b>4 </b>storing probabilities of transitions from any one character to another, and those pieces of the word identification information which correspond to **combinations** of **characters** resulting from the transitions; and an optimization unit for using the character transition probability table a4 in **optimizing** candidate **character strings** obtained by a recognition unit. The word dictionary a<b>6 </b>is searched for a word defined by the **word identification information** which corresponds to the optimized candidate **character string** , whereby the searched **word** is retrieved which applies to the **hierarchy** information and which has yet to be input.

...  
...A character recognition apparatus for inferring the entire **character string** solely from a user-input handwritten keyword and displaying the inferred result as a candidate **character string** . The apparatus of the

invention comprises : a word dictionary storing word identification information and hierarchy information for layering a plurality of words into a hierarchy and for recognizing each of the words within the hierarchy ; a character transition probability table a4 storing probabilities of transitions from any one character to another, and those pieces of the word identification information which correspond to combinations of characters resulting from the transitions; and an optimization unit for using the character transition probability table in optimizing candidate character strings obtained by a recognition unit. The word dictionary is searched for a word defined by the word identification information which corresponds to the optimized candidate character string , whereby the searched word is retrieved which applies to the hierarchy information and which has yet to be input.

...

...A character recognition apparatus for inferring the entire character string solely from a user-input handwritten keyword and displaying the inferred result as a candidate character string .The apparatus of the invention comprises: a word string includes a word hierarchy and for recognizing each of the words within the hierarchy ; a character transition probability table a<b>4 </b> storing probabilities of transitions from any one character to another, and those pieces of the word identification information which correspond to combinations of characters resulting from the transitions; and an optimization unit for using the character transition probability table in optimizing candidate character strings obtained by a recognition unit. The word dictionary is searched for a word defined by the word identification information which corresponds to the optimized candidate character string , whereby the searched word is retrieved which applies to the hierarchy information and which has yet to be input.

**Claims:**

1. A character recognition apparatus having recognition means (a3) for recognizing input character strings and display means (a8) for displaying recognized results, said character recognition apparatus comprising;<br> a word dictionary (a6) storing word identification information and hierarchy information for layering a plurality of words into a hierarchy and for recognizing each of said words within said hierarchy ;<br> a character transition probability table (a4) storing at least probabilities of transitions from any one character to another, and those pieces of said word identification information which correspond to combinations of characters resulting from said transitions ;<br> optimization means (a5) for using said character transition probability table (a4) in optimizing candidate character strings obtained by said recognition means (a3); and <br> retrieval means for searching through said word dictionary (a6) for words defined by those pieces of said word identification information which correspond to the optimized candidate character string , thereby retrieving the searched words which are identified by the applicable pieces of said hierarchy information and which have yet to be input.

...

...die Vorrichtung zum Erkennen von Zeichen umfasst: ein Wörterbuch (a6) zum Speichern von Wortidentifizierungsinformation und Hierarchieinformation zum Einordnen einer Mehrzahl von Worten in eine Hierarchie und zum Erkennen jedes dieser Worte innerhalb der Hierarchie , eine Zeichenumwandlungswahrscheinlichkeitstabelle (a4) zum Speichern von wenigstens Wahrscheinlichkeiten von Umwandlungen eines beliebigen Zeichens

in ein **anderes** und solche Stücke der Wortidentifizierungsinformation, die Kombinationen von Zeichen entspricht, die aus diesen Umwandlungen resultieren...

...Zeichenketten entsprechen, wodurch die gesuchten Worte abgefragt werden, die identifiziert werden durch anwendbare Stücke der **Hierarchieinformation** und die noch eingegeben werdenmüssen.

...

...A character recognition apparatus having recognition means (a3) for recognizing input **character strings** and display means (a8) for displaying recognized results, said character recognition apparatus comprising: a word dictionary (a6) storing word identification information and **hierarchy** information for layering a plurality of words into a **hierarchy** and for recognizing each of said words within said **hierarchy**, a character transition probability table (a4) storing at least probabilities of transitions from any one character to another, and those pieces of said word identification information which correspond to combinations of characters resulting from said transitions; optimization means (a5) for using said character transition probability table (a4) in optimizing candidate **character strings** obtained by said recognition means (a3); and retrieval means for searching through said word dictionary (a6) for words defined by those pieces of said word identification information which correspond to the optimized candidate **character string**, thereby retrieving the searched words which are identified by the applicable pieces of said **hierarchy** information and which have yet to be input.

...

...dictionnaire de mots (a6) stockant des informations d'identification des mots et des informations de **hierarchie** pour structurer une pluralite de mots en une **hierarchie** et pour reconnaître chacun desdits mots au sein de ladite **hierarchie**; un tableau de probabilites de transition des caracteres (a4) stockant au moins des probabilites de transition d'un caractere quelconque a un autre, et les parties desdites informations d'identification des mots qui correspondent a des combinaisons de caracteres resultant desdites **transitions**; un moyen d'optimisation (a5) pour utiliser ledit tableau de probabilites de transition des caracteres...

...d'extraire les mots cherches qui sont identifiées par les parties utilisables desdites informations de **hierarchie** et qui doivent etre entres a present...

... What is claimed is: **1**. A character recognition apparatus having recognition means for recognizing input **character strings** and display means for displaying recognized results, said character recognition apparatus comprising: a word dictionary storing word identification information and **hierarchy** information for layering a plurality of words into a **hierarchy** and for recognizing each of said words within said **hierarchy**; a character transition probability table storing at least probabilities of transitions from any one character to another, and those pieces of said word identification information which correspond to combinations of characters resulting from said transitions; optimization means for using said character transition probability table in optimizing candidate **character strings** obtained by said recognition means; and retrieval means for searching through said word dictionary for words defined by those pieces of said word identification information which correspond to the optimized candidate **character string**, thereby

retrieving the searched words which are identified by the applicable **pieces** of **said** hierarchy information and which have yet to be input.

...

...A character recognition apparatus having recognition means for recognizing input character strings and display means for displaying recognized results, said character recognition apparatus comprising: a word dictionary storing word identification information and hierarchy information for layering a plurality of words into a hierarchy and for recognizing each of said words within said hierarchy; a character transition probability table storing at least probabilities of transitions from any one character to another, and those pieces of said word identification information which correspond to combinations of characters resulting from said transitions; optimization means for using said character transition probability table in optimizing candidate character strings obtained by said recognition means; and retrieval means for searching through said word dictionary for words defined by those pieces of said word identification information which correspond to the optimized candidate character string, thereby retrieving the searched words which are identified by the applicable pieces of said hierarchy information and which have yet to be input.

...

...What is claimed is: 1. A search **system** for searching a multi-item data base, said system comprising: said multi-item data base; a **search**-object specification table for specifying items as search objects; an attribute definition table for specifying...

...a display step of a search result and an indicator indicating how much said search **result** must match a **keyword** or a minimum number of matching searched items in said search result before said search



13/69,K/10 (Item 10 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2007 The Thomson Corporation. All rts. reserv.

0006656771 - Drawing available  
WPI ACC NO: 1994-034595/199404  
XRPX Acc No: N1994-026985

**Document-creating and proofreading support appts. - has input device for character string , proofread data accumulating device connected to editing device with memory using multi-level hierarchical classification, data retrieving device, and proofreading device**

Patent Assignee: SHARP KK (SHAF)

Inventor: KANZA H; KUGA S; MORISHITA T; ONISHI S; WADA M

**Patent Family** (1 patents, 1 countries)

Patent Application

Number	Kind	Date	Number	Kind	Date	Update
US 5280573	A	19940118	US 1990491693	A	19900312	199404 B

Priority Applications (no., kind, date): JP 198962885 A 19890314

#### Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
US 5280573	A	EN	16	10	

#### Alerting Abstract US A

The document-creating support apparatus includes an input device for inputting a **character string** to be proofread, and an editing device functionally connected to the input device for producing an edited sentence. A proofread information accumulating device functionally connected to the editing device includes a memory having retrievable information stored therein according to a multi-level **hierarchical classification**. The retrievable information is stored in records which are linked in the memory by arcs. Each arc has both a pointer to a memory location, an indication of an information type stored at the memory location pointed to by the pointer, and an indication of a level in the multi-level **hierarchical classification**.

A retrieving device is functionally connected to the input device and the proofread information accumulating device, for **specifying a keyword** to be proofread in the **character string**, for retrieving retrievable information related to the keyword from the information accumulating device in accordance with an information type. The proofreading device is functionally connected to the retrieving device for **selecting** retrievable information to replace the **keyword**.

USE/ADVANTAGE - Implements document proofreading and creating functions by utilising information applied to arcs which link correlated retrieving information. Enhanced proofreading accuracy and efficiency.

**Title Terms/Index Terms/Additional Words:** DOCUMENT; SUPPORT; APPARATUS;

INPUT; DEVICE; CHARACTER; STRING; DATA; ACCUMULATE; CONNECT; EDIT; MEMORY ; MULTI; LEVEL; **HIERARCHY** ; CLASSIFY; RETRIEVAL

#### Class Codes

International Classification (Main): G06F-015/62

US Classification, Issued: 395145000, 395155000, 395160000, 395147000, 395146000, 364419140

File Segment: EPI;

DWPI Class: T01

Manual Codes (EPI/S-X): T01-J11A

...has input device for character string , proofread data accumulating

device connected to editing device with memory using multi-level hierarchical classification, data retrieving device, and proofreading device

**Original Titles:**

Document processing support **system** using keywords to **retrieve** explanatory information linked together by correlative arcs

**Alerting Abstract** ...input device for inputting a **character string** to be proofread, and an editing device functionally connected to the input device for producing...

...editing device includes a memory having retrievable information stored therein according to a multi-level **hierarchical** classification. The retrievable information is stored in records which are linked in the memory by...

...pointed to by the pointer, and an indication of a level in the multi-level **hierarchical** classification...

...device is functionally connected to the input device and the proofread information accumulating device, for **specifying** a **keyword** to be proofread in the **character string**, for retrieving retrievable information related to the keyword from the information accumulating device in accordance with an information type. The proofreading device is functionally connected to the retrieving device for **selecting** retrievable information to replace the **keyword**.

**Title Terms...**/Index Terms/Additional Words: **HIERARCHY** ;

**Original Publication Data by Authority**

**Original Abstracts:**

...editing input information. Input information is received from an input means (1) from which a **keyword** is **selected** for **editing** purposes. A low level record is located in a memory (5) which has the keyword of the input information stored therein. The low level record having the **keyword** stored therein is **accessed** to **determine** the location in the memory of higher level records and information types of the higher level records. The higher...

...keyword, a synonym for the keyword, an example usage of the keyword, etc. A display ( 55 ) is provided with an **indication** of the information types stored in the memory for the **keyword** so that a **user** can **select** a desired information **type**. The explanatory information associated with the desired information type is then displayed to the user...

**Claims:**

A document-creating support apparatus, comprising: input means for inputting a **character string** to be proofread; editing means functionally connected to said input means for producing an edited sentence; proofread...

...means and including a memory having retrievable information stored therein according to a multi-level **hierarchical** classification, **said** retrievable information being stored in records which are linked in the memory by arcs, each...

...pointed to by the pointer, and an indication of a level in said multi-level **hierarchical** classification; **retrieving** means functionally

connected to said input means and said proofread information accumulating device for receiving said **character string to proofread**, for **specifying a keyword to be proofread** in said **character string**, for **retrieving** retrievable information related to said keyword from said proofread information accumulating device in accordance with an information type; and proofreading means functionally connected to said retrieving means for **selecting retrievable information** to replace the **keyword**.>

13/69,K/12 (Item 12 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2007 The Thomson Corporation. All rts. reserv.

0004436269 - Drawing available  
WPI ACC NO: 1988-175571/198825

**Computerised information retrieval method for processing text -  
searching text for keywords similar to users search request, selecting  
texts from data base**

Patent Assignee: KLEINBERGER P J (KLEI-I); TNET INC (TNET-N)

Inventor: KLEINBERGER P J; SEAY N J

**Patent Family** (8 patents, 18 countries)

Patent Number	Kind	Date	Application Number	Kind	Date	Update
WO 1988004454	A	19880616	WO 1987US1314	A	19871127	198825 B
AU 198810451	A	19880630				198838 E
DK 198804343	A	19881003				198904 E
EP 334888	A	19891004	EP 1988900194	A	19861202	198940 E
US 4972349	A	19901120	US 1986938163	A	19861204	199049 E
			US 1989393838	A	19890814	
CA 1276728	C	19901120				199101 E
US 5062074	A	19911029	US 1986938163	A	19861204	199146 E
			US 1990575046	A	19900830	
IL 84706	A	19921115	IL 84706	A	19871203	199250 E

Priority Applications (no., kind, date): US 1986938163 A 19861204; US  
1989393838 A 19890814; US 1990575046 A 19900830

#### Patent Details

Number	Kind	Lan	Pg	Dwg	Filing Notes
WO 1988004454	A	EN	68	16	
National Designated States,Original: AU DK FI JP NL					
Regional Designated States,Original: AT BE CH DE FR GB IT LU NL SE					
EP 334888	A	EN			
Regional Designated States,Original: AT BE CH DE FR GB IT LI LU NL SE					
CA 1276728	C	EN			
IL 84706	A	EN			

#### Alerting Abstract WO A

The information retrieval method comprises the steps of defining as a 'criterion key' that key-word which among all the keywords associated with any of the **texts** in the first **group** of **texts**, is associated with the largest number of **texts** within the first **group**. The first group is separated into 2 sub-groups, the first sub-**group** of **texts** having the criterion key as one of its keywords and the second sub-group not including the criterion key.

Results obtained from the above steps are then displayed. The above process is applied **recursively** to at least one of the two sub-groups.

ADVANTAGE - Distinguishes between text areas having sense words but different meanings.

#### Equivalent Alerting Abstract US A

The method uses a processor and associated memory to make explicit the relationships among texts in a text base stored in the memory. The relationships are other than those provided by a user. Each text in the text base of texts associated with keyboard. The method involves the processor accepting from the user a search request of a search to be performed to locate a first **groups** of **texts**. The processor performs the search request described by the user among the keywords associated with the texts in the text base to locate the first **group** of **texts** with associated keywords matching the search request. For each of the keywords

associated with the **texts** in the first **group** , the processor counts the number of texts associated with each of the keywords. The processor compares the number of **texts** the sub- **group** not separated into further sub-groups on the display medium.

(28pp)

#### Equivalent Alerting Abstract US A

The computerised information **retrieval system** is formed of a text base of texts of variable length and content. The texts are selected from the text base on the basis of Boolean logic searches among keywords associated with the **texts** . When a **group** is retrived from such a search, the system automatically segregates the texts based on the presence of absence of a criterion key **keyword selected** so as to segregate the texts into sub-gps. The same criterion key analysis can then be applied **recursively** to the sub-gps. The criterion key analysis can then be applied **recursively** to the sub-gps. The resulting sub-gps. are then displayed to the user in a hierachilical display to illustrate the relationships among the **texts** . A **string** comparison routine is also disclosed to search for similar keywords.

(28pp)

**Title Terms/Index Terms/Additional Words:** COMPUTER; INFORMATION; RETRIEVAL; METHOD; PROCESS; TEXT; SEARCH; KEYWORD; SIMILAR; USER; REQUEST; SELECT; DATA; BASE

#### Class Codes

International Classification (+ Attributes)

IPC + Level Value Position Status Version

G06F-0017/30 A I R 20060101

G06F-0017/30 C I R 20060101

US Classification, Issued: 395144000, 364DIG002, 364927200, 364927400, 364943000, 364943500, 364946200, 364948200, 364948220, 364963000, 364963100, 364974000, 364974300, 364974400, 364974600, 364974300, 364974400, 364974600

File Segment: EPI;

DWPI Class: T01

Manual Codes (EPI/S-X): T01-J05B

**Computerised information retrieval method for processing text...**

...searching text for keywords similar to users search request,  
selecting texts from data base

#### Original Titles:

...INFORMATION **RETRIEVAL SYSTEM** AND METHOD...

...Information **retrieval system** and method...

...Information **retrieval system** and method...

...INFORMATION **RETRIEVAL SYSTEM** AND METHOD

**Alerting Abstract** ...criterion key' that key-word which among all the keywords associated with any of the **texts** in the first **group** of **texts** , is associated with the largest number of **texts** within the first **group** . The first group is separated into 2 sub-groups, the first sub- **group** of **texts** having the criterion key as one of its keywords and the second sub-group not...

...Results obtained from the above steps are then displayed. The above process is applied **recursively** to at least one of the two sub-groups...

**Equivalent Alerting Abstract** ...the user a search request of a search to be performed to locate a first **groups** of **texts**. The processor performs the search request described by the user among the keywords associated with the texts in the text base to locate the first **group** of **texts** with associated keywords matching the search request. For each of the keywords associated with the **texts** in the first **group**, the processor counts the number of texts associated with each of the keywords. The processor compares the number of **texts** the sub-**group** not separated into further sub-groups on the display medium...

...The computerised information **retrieval system** is formed of a text base of texts of variable length and content. The texts...

...the text base on the basis of Boolean logic searches among keywords associated with the **texts**. When a **group** is retrieved from such a search, the system automatically segregates the texts based on the presence of absence of a criterion key **keyword selected** so as to segregate the texts into sub-gps. The same criterion key analysis can then be applied **recursively** to the sub-gps. The criterion key analysis can then be applied **recursively** to the sub-gps. The resulting sub-gps. are then displayed to the user in a hierachilical display to illustrate the relationships among the **texts**. A **string** comparison routine is also disclosed to search for similar keywords...

#### **Original Publication Data by Authority**

##### **Original Abstracts:**

A computerized information **retrieval system** provides a break-down of the major and minor subject areas covered by a **group** of **texts** associated with a **set** of descriptive keywords. The system makes explicit the underlying informational structure of the group as a whole, by organizing the **texts** into sub-**groups** defined by the keywords which the **texts** of each sub-**group** have in common. The process being **recursive**, the sub-groups of each sub-group can be found to any desired depth. An additional method of analysis provides a measure of the degree of similarity between **words** or between **collections** of **words** such as sentences. The method provides a facility for searching for text whose keywords are...

...user's search request and a facility for selecting texts from a textbase and/or **ordering** the lists of **texts** found according to the degree of similarity between the user's search description and textual...

...from the textbase on the basis of Boolean logic searches among keywords associated with the **texts**. When a **group** is retrieved from such a search, the system automatically segregates the texts based on the presence or absence of a criterion key **keyword selected** so as to segregate the **texts** into sub-**groups**. The same criterion key analysis can then be applied **recursively** to the sub-groups. The resulting sub-groups are then displayed to the user in a **hierarchical** display to illustrate the relationships among the **texts**. A **string** comparison routine is also disclosed to search for similar keywords...

...A computerized information **retrieval system** is formed of a textbase of texts of variable length and content. The texts are...

...from the textbase on the basis of Boolean logic searches among keywords associated with the **texts** . When a **group** is retrieved from such a search, the system automatically segregates the texts based on the presence of absence of a criterion key **keyword selected** so as to segregate the **texts** into sub- **groups** . The same criterion key analysis can then be applied **recursively** to the sub-groups. The criterion key analysis can then be applied **recursively** to the sub-groups. The resulting sub-groups are then displayed to the user in a **hierarchical** display to illustrate the relationships among the **texts** . A **string** comparison routine is also disclosed to search for similar keywords...

...A computerized information **retrieval system** provides a break-down of the major and minor subject areas covered by a **group** of **texts** associated with a **set** of descriptive keywords. The system makes explicit the underlying informational structure of the group as a whole, by organizing the **texts** into sub- **groups** defined by the keywords which the **texts** of each sub- **group** have in common. The process being **recursive** , the sub-groups of each sub-group can be found to any desired depth. An additional method of analysis provides a measure of the degree of similarity between **words** or between **collections** of **words** such as sentences. The method provides a facility for searching for text whose keywords are...

...user's search request and a facility for selecting texts from a textbase and/or **ordering** the lists of **texts** found according to the degree of similarity between the user's search description and textual...

13/9/13 (Item 1 from file: 347)  
DIALOG(R)File 347:JAPIO  
(c) 2007 JPO & JAPIO. All rts. reserv.

06920394 \*\*Image available\*\*  
**HIERARCHY** CLASSIFICATION AND RETRIEVAL METHOD

PUB. NO.: 2001-147932 [JP 2001147932 A]  
PUBLISHED: (May 29, 2001) (20010529)  
INVENTOR(s): OKA AKIHIRO  
APPLICANT(s): OKA AKIHIRO  
APPL. NO.: 11-329806 [JP 99329806]  
FILED: November 19, 1999 (19991119)  
INTL CLASS: G06F-017/30

#### ABSTRACT

PROBLEM TO BE SOLVED: To extract a required classification by realizing the AND retrieval of different **words** in high- **order** and low-order classifications even in the case of retrieval with plural words and to extract the classification of a designated rank by designating and retrieving the rank of a classification **hierarchy** with plural words.

SOLUTION: This method comprises a process (a) for inputting plural keywords, a **process** (b) for **retrieving** the titles of respective classifications with one of inputted keywords, a process (c) for preparing a direct high-order classification group including high-order classifications for each of all the classifications of keywords **retrieved** in the **process** (b), a **process** (d) for **retrieving** the titles of respective classifications in the direct high-order classification group prepared in the process (c) with all the other keywords inputted in the process (a), a process (e) for extracting the direct high-order classification group, which includes all the other keywords in the process (d), and defining it as a retrieved answer corresponding to the keyword used in the process (b), a process (f) for finding a retrieved answer corresponding to each of keywords by repeating the processes (b)-(e) concerning the **keywords** except for the **keyword selected** in the process (b), and a process (g) for outputting the retrieved answers prepared in the processes (e) and (f).

COPYRIGHT: (C)2001, JPO



13/9/14 (Item 2 from file: 347)  
DIALOG(R) File 347:JAPIO  
(c) 2007 JPO & JAPIO. All rts. reserv.

02913428 \*\*Image available\*\*  
DOCUMENT RETRIEVING SYSTEM

PUB. NO.: 01-211028 / [JP 1211028 - A]  
PUBLISHED: August 24, 1989 (19890824)  
INVENTOR(s): MORI KEIICHI  
APPLICANT(s): MATSUSHITA GRAPHIC COMMUN SYST INC [330729] (A Japanese  
Company or Corporation), JP (Japan)  
APPL. NO.: 63-035787 [JP 8835787]  
FILED: February 18, 1988 (19880218)  
INTL CLASS: [4] G06F-007/28  
JAPIO CLASS: 45.1 (INFORMATION PROCESSING -- Arithmetic Sequence Units);  
45.2 (INFORMATION PROCESSING -- Memory Units)  
JOURNAL: Section: P, Section No. 962, Vol. 13, No. 517, Pg. 153,  
November 20, 1989 (19891120)

#### ABSTRACT

PURPOSE: To ensure the integrated registration of key words, the sure retrieval of documents, and the quick comparison by **selecting the key word** items turned into a **hierarchical** form by an interactive process and **designating** the key **words in combinations**.

CONSTITUTION: The most significant **hierarchy** of a **group** of key word items stored in a storage part is displayed and the item group of the next higher **hierarchy** related to one item if selected by an input device 5 is displayed. Such an interactive process is carried out down to the item group of the lowest **hierarchy**. A key word code (K.C.) consisting of the item numbers of all selected **hierarchies** is supplied to a retrieving part. Then an input K.C. is stored in a key word storing part 9 corresponding to the document retrieving information designated by a control part 2 in a registration mode. In a document retrieving mode the document retrieving information coincident with the input K.C. is read out of the part 9 and given to the part 2. Then the document retrieving information is designated at a retrieving part 10 in case of registration and the corresponding document is detected out of a document storing part 1 based on the document retrieving information in case of the document retrieval.

13/9/15 (Item 3 from file: 347)  
DIALOG(R)File 347:JAPIO  
(c) 2007 JPO & JAPIO. All rts. reserv.

02608957 \*\*Image available\*\*  
RESEMBLING CHARACTER RETRIEVAL SYSTEM

PUB. NO.: 63-225857 [JP 63225857 A]  
PUBLISHED: September 20, 1988 (19880920)  
INVENTOR(s): OTA TATSUO  
MATSUMOTO SHUNJI  
APPLICANT(s): FUJITSU LTD [000522] (A Japanese Company or Corporation), JP  
(Japan)  
APPL. NO.: 62-057957 [JP 8757957]  
FILED: March 14, 1987 (19870314)  
INTL CLASS: [4] G06F-015/20; G06F-007/28  
JAPIO CLASS: 45.4 (INFORMATION PROCESSING -- Computer Applications); 45.1  
(INFORMATION PROCESSING -- Arithmetic Sequence Units); 45.2  
(INFORMATION PROCESSING -- Memory Units)  
JOURNAL: Section: P, Section No. 815, Vol. 13, No. 27, Pg. 84, January  
20, 1989 (19890120)

#### ABSTRACT

PURPOSE: To facilitate operation by detecting the item, which most resembles a keyword segmented from an input **character string**, in the range from a noticed item to items a prescribed level lower than this item out of items **hierarchically** provided in the data base to retrieve a desired item.

CONSTITUTION: A reading part 4, a resemblance degree calculating part 5, and an item specifying part 6 are provided. The reading part 4 successively reads out all items in the range from a noticed item to items in the **hierarchy** a prescribed level lower than item in the data base as candidates corresponding to the keyword segmented from the input **character string**, and the resemblance degree calculating part 5 calculates the degree of resemblance between each item read by the reading part 4 and the **keyword**, and the item **specifying** part 6 specifies an item based on the degrees of resemblance calculated by the resemblance degree calculating part 5. The item specified by the item specifying part 6 is noticed, and all items in the range from this item to items in the **hierarchy** the prescribed level lower than this item are read out, and degrees of resemblance between these items and the next **keyword** are calculated to **specify** an item, and this operation is repeatedly executed to retrieve a desired item. Thus, the operation for item selection is facilitated.